Essays on Human Capital and Financial Economics

by

Jialan Wang

B.S., Mathematics, California Institute of Technology, 2003

Submitted to the Alfred P. Sloan School of Management
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2010

Author ....
Alfred P. Sloan School of Management
May 17, 2010

Certified by ....
Antoinette Schoar
Michael Koerner '49 Professor of Entrepreneurial Finance
Thesis Supervisor

Accepted by ....
Ezra W. Zuckerman
Director, Ph.D. Program, Sloan School of Management

Essays on Human Capital and Financial Economics

by

Jialan Wang

Submitted to the Alfred P. Sloan School of Management
on May 17, 2010, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis consists of three essays examining issues related to human capital, careers, and financial economics. In the first chapter, I examine how the process of corporate bankruptcy varies by human capital intensity using a sample of 1,493 public firms that filed for Chapter 11 between 1980 and 2003. I document two key patterns. First, human-capital-intensive are more likely to avoid and delay bankruptcy conditional upon entering distress, and they are more likely to use debt issuance to raise funds prior to bankruptcy. Second, human-capital-intensive firms are more likely to be liquidated within bankruptcy.

In the second chapter (co-authored with Pierre Azoulay and Joshua Graff Zivin), we estimate the magnitude of human capital spillovers generated by 112 academic "superstars" who died prematurely and unexpectedly, thus providing an exogenous source of variation in the structure of their collaborators' coauthorship networks. Following the death of a superstar, we find that collaborators experience, on average, a lasting 5 to 8% decline in their quality-adjusted publication rates. By exploring interactions of the treatment effect with a variety of star, coauthor and star/coauthor dyad characteristics, we find evidence that spillovers are circumscribed in idea space, but less so in physical or social space. In particular, superstar extinction reveals the boundaries of the scientific field to which the star contributes — the "invisible college."

In the third chapter, I examine the role of artistic films in the careers of star actors and directors. Using data from all films released in the United States from 1980 and 2005 and the career histories of 100 star directors and 94 star actors, I document evidence on the interaction between artistic films and the value of stars over their careers. Artistic films make up 12% of star careers, and they are associated both with significantly lower film revenues and lower monetary compensation. The propensity for stars to work on artistic films is relatively constant across their career, although it is slightly higher when stars are under 30 or over 60 relative to middle age. Furthermore, artistic films are significantly associated with Oscar awards.

Thesis Supervisor: Antoinette Schoar
Title: Michael Koerner '49 Professor of Entrepreneurial Finance

Acknowledgments

Above all else, my thesis is a testament to how lucky I have been to receive the support of so many family, friends, advisors, and mentors throughout my life. First of all, I would like to thank my husband Patrick LaBlance for his unwavering love and moral support during the past six years of graduate school and eight years of marriage, for believing in me when I was mired in uncertainty and for pushing me when I didn't feel I could go forward. I owe an irreparable debt to my parents Yingni Che and Xiaoshen Wang, who have never failed me and have stood behind all of my endeavors.

I have reached this milestone in my intellectual journey only by standing on the shoulders of the succession of scholars and mentors who nurtured my creativity and confidence. To Chris Adami, Stan Franklin, Alan Hajek, Fred Hickling, Henry Lester, and Norb Schedler, thank you for showing me the way. I could not have sustained my passion for the academic endeavor without you.

My committee members Pierre Azoulay, Carola Frydman, Gustavo Manso, and Antoinette Schoar have been a crucial source of inspiration and guidance during my doctoral studies. In addition, I have received a great deal of help and advice from other faculty including David Autor, Serdar Dinç, Ken Froot, Dirk Jenter, Josh Lerner, Hamid Mehran, Stew Myers, Drazen Prelec, and Steve Ross. I gratefully acknowledge financial support in 2007 from the State Farm Companies, Inc.

I have benefitted greatly from my interactions with fellow graduate students, including Manuel Adelino, Tara Bhandari, David Cesarini, Chuck Easley, Alex Edmans, Dimitris Papanikolaou, Fadi Kanaan, Cynthia Kinnan, Jiro Kondo, Lynn Li, Mary Tian, Ngoc-Khanh Tran, Ding Wu, Lynn Wu, Yifan Zhou, and most of all, Soojin Yim and the inimitable Jack Bao. I am also grateful to Eric Beaton, Beth Conlin, Peter Cooper, Jennifer DiMase, Amanda Graham, Stephen Intille, Steve Lanou, and the members of Closing the Loop for helping me give back to the MIT community with the little time and energy I could steal from my studies.

Finally, I'd like to thank my many dear friends for making my PhD a journey of personal, intellectual, and moral growth. Thanks to Ayeh Bandeh-Ahmadi, Saskya Byerly, Juhi Chandalia, Joy Chien, Sherri Goings, Dana Ionita, Dave McKeen, Vikram Mittal, Nina Wickens, Robert Wang, and Virgi Vassilevska Williams for always being there for me. To Zoe Chance, Jason Jay, Kendra Johnson, Weiji Ma, Yoda Patta, Eliza Petrow, Ari Shapiro, and Cherry Wan - you have made this all worthwhile by showing me the meaning of giving and empowering me to make a positive impact on our world.

Contents

List of Figures

List of Tables

Chapter 1

The Role of Human Capital in Corporate Bankruptcy

1.1 Introduction

Bankruptcy and the resolution of financial distress have long been central topics in corporate finance (Altman [1968]), and a substantial empirical literature has documented the path of financial distress for firms that file for Chapter 11.[1] A number of researchers have recognized that the nature of a firm's assets plays an important role in its experience of distress (Shleifer and Vishny [1992], Pulvino [1999], Acharya et al. [2007]). Despite the growing importance of human capital for firms in the economy, however, there has been little empirical work addressing its specific role in the bankruptcy process.[2] The inalienability of human capital may be especially salient in the context of bankruptcy, when the relationships between firms and their employees become threatened. Indeed, bankrupt firms routinely cite employee retention as a critical concern.[3] Because of their vulnerability to employee flight, human-capital-intensive firms are likely to suffer greater costs of financial distress.

In this study, I consider whether the behavior of human-capital-intensive firms before, during,

[1] Asquith et al. [1994] examine firms' responses to distress prior to entering bankruptcy, which often include asset sales, debt restructuring, and cuts in capital expenditure. Gilson et al. [1990] compare successful out-of-court restructurings and bankruptcy filings. Kalay et al. [2007], Denis and Rodgers [2007], and Lemmon et al. [2009] consider the determinants of firm outcomes within bankruptcy, and Hotchkiss [1995] and Lemmon et al. [2009] document post-bankruptcy performance.

[2] Berk et al. [2007], Berkovitch et al. [1997], Jaggia and Thakor [1994] are notable theoretical papers which relate human capital and bankruptcy.

[3] Firms often submit motions during the first day of bankruptcy proceedings appealing the court to authorize timely payment of wages outstanding at the time of filing, which would otherwise remain unpaid until the conclusion of the Chapter 11 case. As a typical example, a motion filed in the case of Interliant, Inc. in 2002 asserted that "The Debtors' employees are central to their operations and are vital to their reorganization. A deterioration in personnel morale at this critical time undoubtedly would have a devastating impact on the workplace, the value of the Debtors' assets and business and its ability to benefit from Chapter 11."

and upon emergence from Chapter 11 is consistent with their experience of higher costs of financial distress. I highlight two key patterns. First, conditional on entering financial distress, human-capital-intensive firms are more likely to avoid and delay entry into Chapter 11. Human-capital-intensive firms are also more likely to issue debt in the years prior to bankruptcy, resulting in a faster increase in leverage. The second major finding is that human-capital-intensive firms are more likely to be liquidated within bankruptcy, suggesting. Moreover, they perform better conditional upon emergence from Chapter 11, suggesting that their greater liquidation rates are not driven solely by the differential selection of firms into bankruptcy. These findings have implications for both the capital structure decisions of human-capital-intensive firms and the effectiveness of asset reallocation in bankruptcy.

In the absence of firm-level data on human capital, I follow the labor economics literature in using the Current Population Survey (CPS) to measure human capital intensity. My main measure is the share of college-educated workers in each industry (college share).[4] Although college share seems to capture an intuitive notion of human capital intensity (see Table 1.3), the results are consistent when using a variety of specifications and alternative human capital measures.

First, I examine the selective entry of firms into bankruptcy among publicly-listed firms that enter financial distress. Consistent with the avoidance of bankruptcy in anticipation of greater costs of financial distress, human-capital-intensive firms that enter distress are more likely to recover from distress or be acquired by other firms and less likely to enter Chapter 11. Nonetheless, a significant fraction of distressed firms ultimately end up in Chapter 11, and the remaining results pertain to 1493 Chapter 11 filings by publicly-listed firms between 1980 and 2003.

While the previous literature has emphasized restructuring activities such as selling assets and reducing investment that firms undertake in response to financial distress (Asquith et al. [1994], Brown et al. [1994]), I find that human-capital-intensive firms are less likely to take measures that reduce firm scale. In particular, they are less likely to reduce their asset size, investment, and employment in response to distress prior to entering bankruptcy. Instead, human-capital-intensive firms cope with earnings shortfalls by increasing borrowing levels.

Next, I analyze how human capital relates to firm outcomes upon entering bankruptcy. Firms with above-median college share are nine percent more likely to be liquidated within bankruptcy compared with low-college-share firms. Among firms that emerge and remain publicly listed, however, those with above-median college share are fourteen percent *less* likely to refile for bankruptcy within five years of emergence. These results suggest that the greater liquidation rates of human-capital-intensive firms are not driven solely by differential selection into bankruptcy.

[4] See Section 1.3 for details on variable construction.

1.2 Sample description

The sample consists of publicly-held firms which enter Chapter 11 between 1980 and 2003. This sample period is chosen so that firms fall under the 1978 Bankruptcy Reform Act, which marked the most recent drastic change to the U. S. bankruptcy law.[5] To accomplish this task, I match the Compustat database with bankruptcy filings since January 1, 1980 in WebBRD (Bankruptcy Research Database),[6] the Capital Changes Reporter, and New Generation Research.[7] To minimize truncation bias for the most recent cases, I restrict the sample to firms which file before December 30, 2003. Consistent with the literature, I exclude firms in the finance and utilities industries (two-digit NAICS codes 22 and 52) due to differences in regulatory oversight and accounting standards for these industries. This data collection process yields a sample containing 1493 Chapter 11 filings by 1405 firms.

For each firm in the sample, I collect data related to three distinct phases of the bankruptcy process: firm characteristics prior to entering bankruptcy, firm outcomes at the resolution of bankruptcy, and post-bankruptcy performance and outcomes for firms which emerge and remain publicly listed. Financial accounting data are collected from Compustat and 10-Ks beginning five years before bankruptcy filing until five years after emergence (for firms which emerge), up to 2008. Annual stock returns are collected from CRSP. Following Asquith et al. [1994] and Kalay et al. [2007], I measure operating performance using *profitability*, defined as earnings before interest, depreciation, and amortization (EBITDA) divided by the book value of assets. Although earnings may be subject to manipulation and smoothing by managers, they are likely to be a reasonable proxy for cross-sectional differences in operating performance. In most of the analysis, I adjust profitability by subtracting its contemporaneous industry median among Compustat firms, which mitigates differences in accounting practices and other sources of industry-level variation in profitability. Furthermore, the results are robust to measuring performance using operating income instead of EBITDA. Another possible problem with the *profitability* measure is that the book value of assets is likely to underestimate the value of human capital and intangible assets. However, the results are robust to normalizing EBITDA by total sales instead of book value, so this bias does not appear to drive the results.

To facilitate comparison to prior literature, I present two measures of leverage: the book value

[5]The 1978 Bankruptcy Reform Act established the current system of federal bankruptcy courts and the regime of Chapter 11 reorganization, and became effective on October 1, 1979. More recently, the Bankruptcy Abuse Prevention and Consumer Protection Act of 2005 enacted moderate changes to the 1978 law, but the effects of these changes are excluded from my analysis.

[6]http://lopucki.law.ucla.edu/bankruptcy_research.asp, Accessed February 2008. The WebBRD covers firms with at least $100 million in assets which filed for bankruptcy starting in 1980. I am grateful to Lynn Lopucki for sharing his dataset for use in this paper.

[7]From bankruptcydata.com, accessed April 2009

of liabilities divided by the book value of assets, and the book value of current and long-term debt divided by the book value of assets.[8] Book measures of leverage are more appropriate for capturing the borrowing behavior of distressed firms than market measures, since large declines in the market value of equity prior to bankruptcy would drive up the standard measure of market leverage (book value of debt / (book debt + market equity)) even when the underlying level of liabilities has not changed. Moreover, market values of debt are generally unavailable. In addition to profitability and leverage, I also collect data on employment, interest expenses, property, plants and equipment (PPE), and research and development (R&D) from Compustat and 10-Ks, and I collect data on equity values and returns from CRSP.

I gather information on the outcomes of the Chapter 11 cases from WebBRD, New Generation Research, court documents [9], 8-Ks, and news searches. Similarly to previous studies (Hotchkiss [1995], Kalay et al. [2007], Bharath et al. [2007]), I sort firm outcomes at resolution into four categories: *Emerged public, Emerged private, Liquidated*, and *Acquired*. *Emerged* firms are those which successfully confirm a plan of reorganization and exit bankruptcy as independent operating companies, and the *Emerged public* subsample consists of firms which subsequently appear in the Compustat database, while *Emerged private* indicates firms which were not subsequently listed in Compustat. Firms are classified as *Liquidated* if substantially all of their assets were sold in a piecemeal fashion, which can occur either within Chapter 11 or through a conversion to Chapter 7. *Acquired* firms are those which sold the majority of their assets as a going concern to a single purchaser. While the full sample is used for the analysis of firm behavior and outcomes prior to and during bankruptcy, the analysis of post-bankruptcy outcomes is restricted to the *Emerged public* sample due to data availability and to ease comparisons to the literature.

Overall, the firms in my sample are larger but in similar financial condition compared with most samples studied previously.[10] Table 1.1 presents selected summary statistics for the bankruptcy sample corresponding to the last fiscal year-end prior to filing, splitting the sample by firm outcome at the resolution of bankruptcy. The average (median) firm in the sample has revenues of \$614 million (\$126 million), which is substantially larger than those in most previous studies.[12]

[8]As discussed by Welch [2008], debt / assets may mismeasure the true variation in financial leverage across firms. Moreover, creditors of both financial and non-financial liabilities are important claimants in the process of bankruptcy, so in most of the analysis I use total liabilities divided by assets. Nonetheless, the results of the paper are insensitive to choice of leverage measure.

[9]Case dockets and documents were obtained cases starting in the 1990s via the Public Access to Court Electronic Records (PACER) system.

[10]However, one important caveat is that firms in my sample experience relatively less severe distress upon entering Chapter 11 compared with those studied by Hotchkiss [1995], who reports substantially higher financial leverage and more negative operating performance [11]. Because much of Hotchkiss [1995]'s sample is drawn from non-Compustat firms, these comparisons indicate that sample selection into Compustat introduces substantial differences in observable firm characteristics which should be taken into account when comparing our results.

[12]For example, Hotchkiss [1995] reports a mean sales of \$ 260 million for her full sample. When I restrict my sample to match her sample period, mean sales is still twice as large as that in her sample. This discrepancy results

Unsurprisingly, the sample is quite highly-leveraged, with mean (median) liabilities over assets of 1.19 (0.88), and debt over assets of 0.71 (0.55). With the exception of Hotchkiss [1995] who reports liabilities over assets of 2.50, the firms in my sample are similarly leveraged compared with those in previous studies.[13] Gilson et al. [1990] report mean liabilities over assets of 1.01, and Weiss [1990] reports mean debt over book assets of 0.77. The mean (median) interest coverage in my sample is -7.56 (0.04), indicating that the majority of firms in the sample have serious cashflow shortfalls, and some experience extreme shortfalls due to very low earnings. The average profitability for the sample is -0.25 with a median of 0.001, indicating that these firms suffer from quite severe operating difficulties in addition to pure financial distress stemming from high leverage.

Among the four bankruptcy outcomes, nearly half of the firms emerge from bankruptcy as independent entities, with 30% emerging under public ownership and 18% under private ownership. Most of the remaining (38%) are liquidated in bankruptcy, while 11% are acquired. Outcomes could not be ascertained for the final 4% of firms. Consistent with Lemmon et al. [2009] and Denis and Rodgers [2007], pre-bankruptcy characteristics differ strongly between firms which are ultimately liquidated and acquired in bankruptcy and those which emerge.

Comparing the characteristics of firms with different outcomes reveals a preliminary picture of the selection process within bankruptcy. Firms which are liquidated tend to be smaller, have lower profitability, and have lower leverage prior to bankruptcy than those which emerge, and most of these differences are significant at the 1% level for both means and medians. Acquired firms tend to be smaller and have lower profitability than those of the other two groups, while their leverage falls between that of emerged and liquidated firms. The duration of Chapter 11 is substantial, lasting an average of 19 months in the full sample. Furthermore, firms which are liquidated spend significantly longer in bankruptcy than those which emerge, with mean (median) durations of 22 (16) months versus 17 (13) months, respectively.

---

from differences in data construction, as much of her sample is not listed in Compustat (her sample covers 806 bankruptcies while mine includes only 171 filings during the same period). When I restrict my sample to match the period examined by Gilson et al. [1990], firm size matches much more closely. Gilson et al. [1990] report a has a mean book value of $317 million and median of $49 million for 61 firms which file for bankruptcy, while my subsample of 92 firms has a mean book value of $270 million with median of $80 million. Thus, the discrepancies in firm size between my sample and previous ones results both from secular increases in firm size and from sample selection into Compustat.

[13]These differences persist when restricting my sample to the same sample period (1978 to 1989) as Hotchkiss [1995] or restricting to firms with less than $400 million in total assets to obtain similar means in firm size. Thus, the differences in financial condition between my sample and that of Hotchkiss [1995] appears to be driven by the inclusion of non-Compustat firms in her sample.

1.3 Human capital measures and industry composition

Because publicly-available data sources do not provide consistent firm-level data on human capital for my sample,[14] I measure human capital at the industry level using data from the Current Population Survey (CPS).[15] Conceptually, a firm's human capital encompasses all of the characteristics of its employees that contribute to its productive output, including schooling, training, health, and innate ability. Among the many possible characteristics related to human capital, schooling has been the main focus of human capital research since the seminal contributions of Becker [1962] and Mincer [1974], and it is a natural place to start for constructing a consistent measure which can be applied to a wide range of firms. To summarize levels of worker schooling, I focus on the share of college-educated workers in the labor force, which is denoted by *college share* in the rest of the analysis. I emphasize college education because trends in wage growth and productivity suggest that college-educated workers have been the primary drivers of economic growth in the past few decades[Jorgenson et al., 2003], and much of the work relating finance and human capital has highlighted the role of knowledge workers in the 21st-century firm [Zingales, 2000].[16] I define *college share* for each industry in each year between 1980 and 2003 as the share of full-time, employed workers with at least 16 years of education.[17]

The college attainment of the U. S. workforce has increased steadily in the last several decades. Based on the CPS sample, the average college share among full-time workers in the private sector has increased dramatically from 15.3% in 1980 to 25.4% in 2003, as shown in Figure 1-1. Furthermore, these trends are driven largely by within-industry increases in education level. But does the secular increase in college share reflect real changes in the importance of human capital, or simply demographic changes as college education has become more widely-available? Since both the supply of college-educated workers and the returns to college education rose during this period, supply changes alone cannot account for the rising college premium(Katz and Murphy [1992]). Furthermore, many researchers suggest that the information technology revolution has increased the returns to the cognitive and analytical skills of college graduates[Autor et al., 2003]. Because the drivers of cross-sectional versus time-series variation in college attainment are likely to differ substantially, I consider both time-varying and cross-sectional measures of college share which will be described in more detail below.

Although I focus on schooling, it is important to note that there are many other facets of human

[14]While Compustat reports total numbers of employees, staff expenses, and selling, general, and administrative expenses, the latter variables suffer from low data coverage and encompass a wide range of employee-related expenses, with accounting standards varying across firms.

[15]The CPS is a monthly survey conducted by the Bureau of the Census consisting of about 100,000 individuals across the population.

[16]My results are qualitatively similar if I use the number of years of educational attainment.

[17]See the Data Appendix for details on the construction of this measure

capital including experience, training, and personal attributes such as innate ability, motivation, and health. Work experience has been a major element of human capital literature since the work of Ben-Porath [1967] and Mincer [1974], but it is problematic for my purposes because experience is highly collinear with age, which may confound its interpretation. Training, often sponsored by firms themselves, is another significant component of human capital (Acemoglu and Pischke [1999], Autor [2001]), but a lack of comprehensive and standardized data on training across firms or industries make it difficult to incorporate in my setting (Bassi et al. [2000]). Finally, data limitations hamper the ability to account for other components of human capital such as school quality and innate ability.

While my main results rely on the college share measure to proxy for the relative importance of human capital across industries, I attempt to distinguish between general and firm-specific human capital in an extension to my results in Section 1.7. Because of the inherent difficulties in separating these two components of human capital and distinguishing firm-specific human capital from other characteristics of workers, firms, and industries, the measure I present should be interpreted with caution. With this caveat in mind, I proceed to describe a simple proxy for firm-specific human capital.

Conceptually, a straightforward view is that firm-specific human capital consists of the skills which arise from the accumulation of on-the-job experience and from investments made by the firm and employee (e.g. employer-sponsored training, effort spent learning firm procedures) to specialize the employee's skills to match the firm's needs [Becker, 1964]. A rather different view asserts that firm-specific human capital arises from matching, in which information about the suitability of a worker for a particular firm is gradually revealed over time Jovanovic [1979]. However, neither the extent of investment in firm-specific skills nor the accumulation of knowledge about match quality can generally be observed, making measurement of firm-specific-human-capital challenging.

Nonetheless, under both views the presence of firm-specific human capital generates a surplus as long as the firm and employee remain attached. The value of this surplus provides an incentive for maintaining the employment relationship, thus leading to higher observed tenure in firms where firm-specific human capital is important. The virtue of using employer tenure as a proxy for firm-specific human capital is that higher tenure (and its corollary, lower turnover) is an outcome shared among the various mechanisms for attaining firm-specific human capital,[18] so its validity does not rest on any particular model or channel. To make sure both the firm and employee have an incentive to maintain their relationship, most of these models also predict that firms and employees share the surplus generated by firm-specific human capital,[19] so a test of tenure as a proxy for firm-specific

[18]See Parsons [1972], Hashimoto [1981], and Jovanovic [1979] for theoretical models which predict the relationship between firm-specific human capital and lower turnover.

[19]Becker [1964] advanced the intuitive notion that firms and employees should share both the costs of accumulating

human capital is whether wages tend to rise with employer tenure. Confirming this prediction, the positive relationship between wages and tenure has been well-documented empirically.[20] To construct an empirical proxy for firm-specific human capital from the CPS, I define *employer tenure* for each industry in each year as the average tenure among full-time, employed workers in the private sector(see Appendix Section 1.10.3 for details).

An obvious weakness of the employer tenure measure is that workers can have high tenure for reasons unrelated to firm-specific human capital. In particular, some industries could be associated with higher rents or different institutions, often leading both to higher tenure and higher wages (Katz and Summers [1989]). For example, tenure tends to be high in the public sector, although it is not obvious that firm-specific human capital is more important for government workers. However, a substantial literature on displaced workers (most notably Jacobson et al. [1993]) shows that earnings declines for workers displaced from their employers due to involuntary layoffs are significantly correlated with worker tenure, a result which is inconsistent with a simple rent-based explanation. These results provide evidence both that tenure is correlated with productive value and that this source of human capital is degraded when workers are separated from the original firm. Another possible confound is that labor unions provide to more bargaining power for workers in addition to greater job security. Indeed, union members have an average of 12.5 years of tenure compared with 6.8 for non-union members. However, in my analysis I test for the robustness of my results to controls for industry unionization, and find that the average percentage of unionized workers in an industry is generally insignificantly related to bankruptcy outcomes, and its inclusion does not affect my results.

Between 1981 and 2002, employer tenure increased from 6.3 years to just over 8 years by 1987 and subsequently declined to 6.6 years by 2002 (Figure 1-1). The secular trend in employer tenure may be caused by three main types of forces: demographic changes, business cycle effects, and changes in the importance of firm-specific human capital. Because tenure tends to rise with a worker's age, the initial rise in tenure may be partially attributed to the aging of the baby boomer generation. Furthermore, the increased labor force attachment of women also contributed to rising tenure levels. The decline in tenure levels since the late 1980s has been well-documented, and can largely be attributed to job loss in high-tenure industries such as the manufacturing sector (Farber [2007]). Since the tenure measure I use is based on workers who remain employed and low-tenure workers are more likely to be laid off during downturns, tenure levels are counter-cyclical. From this discussion, utilizing time-series variation in employer tenure may be problematic for measuring firm-specific human capital due to its sensitivity to demographic and business cycle effects.

---

firm-specific human capital and the surplus it generates, while Hashimoto [1981] first formalized the sharing of surplus.

[20]See for example Topel [1991]

In summary, while college share and employer tenure are imperfect proxies for human capital, they satisfy several key criteria. Both are consistently measured across time and across industries, and both measures are positively related to wages, indicating that they are correlated with the value of employees' services to their firms. Furthermore, the wage premia arising from employer tenure are diminished when workers are displaced involuntarily from their jobs, consistent with the role of tenure as a proxy for firm-specific skills.

Because both proxies can be influenced by factors unrelated to human capital which vary across time and across industries, I consider several variations of *college share* and *employer tenure* which differ in their exploitation of time series and cross-sectional variation: 1) panel measures which are calculated for each industry and year, 2) cross-sectional averages within each industry, and 3) cross-sectional levels in 1980. For each industry i, year t, the variations for *college share* are defined as

$$CollegeShare_{i,t}^{panel} = \frac{\sum_{j\in i} w_{j,t} college_{j,t}}{\sum_{j\in i} w_{j,t}}$$

$$CollegeShare_{i}^{average} = \frac{\sum_{t=1980}^{T=2003} CollegeShare_{i,t}}{24}$$

$$CollegeShare_{i}^{initial} = \frac{\sum_{j\in i} w_{i,1980} college_{j,1980}}{\sum_{j\in i} w_{j,1980}}$$

where $w_{j,t}$ is the CPS sampling weight and $college_{j,t}$ is a dummy variable for college education for individual j. These measures are defined similarly for *employer tenure*.

The panel measures may be problematic because of business-cycle fluctuations and well-documented secular changes in demographics, which particularly affect employer tenure. As long as these demographic trends affect all industries, using year fixed-effects would help mitigate their effects when using the panel measures. Other ways to capture cross-sectional variation are by taking the average of college share and employer tenure within each industry across the sample period and by taking their initial levels at the beginning of the sample period. Cross-sectional rankings of industries by both college share and employer tenure have also been persistent over time[21], so the $average$ and $initial$ measures represent good summaries of cross-sectional variation. Due to the descriptive nature of this study, I will keep the source of variation as transparent as possible by focusing on results using the $average$ measures, although the results are robust to using the $panel$ and $initial$ measures[22].

Table 1.3 describes the industry composition for the bankruptcy sample along with the $average$

[21]The spearman rank correlations across three-digit industries are 0.68 between 1980 and 2003 for college share and 0.78 between 1981 and 2002 for employer tenure.

[22]Results are robust when using the *panel* measures with either year or industry fixed-effects.

measures of college share and employer tenure. For brevity, industries are collapsed to two-digit NAICS groups, although the human capital variables are measured at the three-digit level in the analysis. The bankrupt firms come from 18 distinct two-digit NAICS groups, with the largest number of filings occurring in the manufacturing and retail sectors. Industry shares are consistent with those reported in previous studies [23] as well as the Compustat universe, although retail is slightly over-represented within the bankruptcy sample and manufacturing is under-represented. College share ranges from 7 percent in agriculture to 55 percent in educational services, while average employer tenure ranges from 4.1 years in accommodation and food services to 9.5 years in wood product manufacturing.

Table 1.4 displays the correlations between the cross-sectional human capital measures and firm characteristics for the bankruptcy sample which have been averaged within the 70 three-digit NAICS industries in the bankruptcy sample. Firm characteristics are equally-weighted and measured at the fiscal year-end before filing. Q is defined as (book value of assets + market value of common equity - book value of equity - book value of deferred taxes) / book value of assets, *tangibility* is defined as net property, plants and equipment (PPE) / book value of assets, and *R&D/book* is R&D expenses / total assets. *Profitability* is as defined above, and in this table and the remaining analysis I use the definition of *Leverage* based on total liabilities / book value. College share and employer tenure are negatively correlated with a correlation coefficient of -0.12.[24] College share is positively correlated with Q and R&D intensity, while it is negatively correlated with log sales, leverage, tangibility, and profitability. Employer tenure exhibits precisely the opposite pattern of correlation.

Overall, Tables 1.3 and 1.4 show that firms with high college share tend to be found in the high-technology and skilled service sectors, and they are characterized by higher growth opportunities and R&D, and low levels of asset tangibility. In contrast, firms with high levels of employer tenure human capital tilt heavily toward manufacturing, and is associated with more mature businesses that have high profitability, lower growth opportunities, and higher asset tangibility. These classifications fit with intuitive notions of the kinds of firms which require high levels of human capital and more-specific human capital. While professional service and high-technology firms need workers with the advanced cognitive skills associated with high levels of overall human capital, manufacturing firms rely on complex production systems which are largely firm-specific and may require extensive on-the-job training to maximize employee productivity.

Industries which differ by human capital are also different on a variety of other dimensions, so it is important to take into account potential biases from omitted variables in my results. I address this

[23] See Hotchkiss [1995] and Gilson [1990]

[24] The correlation between tenure and college education at the worker level in the CPS supplements which report employer tenure is -0.02.

issue explicitly in Section 1.8. A remaining concern is that my measures do not accurately capture the level of human capital in the firms in my bankruptcy sample. As documented by Abowd et al. [2005], there is substantial heterogeneity in the human capital content of firms within an industry. Thus, human capital levels for individual firms are measured with error by my industry-level proxies. Due to lack of firm-level data on human capital, however, further investigation of these concerns is left to future research.

1.4 Human capital and pre-bankruptcy behavior

If human-capital-intensive firms experience greater bankruptcy costs, then they should be less likely to enter Chapter 11 conditional upon experiencing financial distress, due both to changes in operating policy by the firm and to renegotiation with creditors. As documented by Asquith et al. [1994], firms typically attempt to avoid bankruptcy by engaging in asset sales, out-of-court restructuring, and reductions in capital expenditures. In this section I explore whether human-capital-intensive firms are able to avoid bankruptcy to a greater extent and the tactics they use to do so.

I define the onset of financial distress among Compustat firms as in Asquith et al. [1994], as the first of at least two consecutive years of interest coverage (EBITDA / interest expense) below one, or any year in which interest coverage is less than 0.8 [25], and I identify 19,053 firm-years in which firms experience the onset of financial distress in the overall Compustat sample. Panel A of Table 1.2 shows firm outcomes 10 years after the initial onset of financial distress as defined by Compustat exit codes, where the initial distressed sample is split by the median college share. The Table presents preliminary evidence that human-capital-intensive firms anticipate higher bankruptcy costs. Conditional upon experiencing financial distress, high-college-share firms are 5% more likely to merge with another firm, 4% less likely to enter bankruptcy, and 4% more likely to survive as an independent entity compare with low-college-share firms. Thus, human-capital-intensive firms appear better able to avoid bankruptcy by arranging a merger or renegotiating with creditors. Nonetheless, a substantial fraction do enter bankruptcy, indicating that frictions in the renegotiation process and in the market for corporate control limit the ability of human-capital-intensive firms to respond to distress. In the remaining analysis, I focus only on the sample of firms that eventually file for Chapter 11.

As a first indication of differences in the extent to which firms respond to distress prior to entering bankruptcy, I consider in Panel B of Table 1.2 how human capital relates to the length of time between initial distress and bankruptcy. The first row compares the number of years between

[25] If financial distress continues for more than one year, I consider only the first year of any continuous period of distress.

the end of the first fiscal year of financial distress and the date of bankruptcy filing, where the onset of distress is defined as above. The mean (median) number of years of distress prior to bankruptcy is 3.72 (2.99) for low-college-share firms, compared with 4.41 (4.08) for high-college-share firms, and the differences are statistically significant at the 1% level. One potential issue with the definition of distress onset in specification (1) is that industries with different levels of human capital differ in their average levels of interest coverage[26] or in the lengths of their business cycles. Thus, in the second row, I define the onset of distress as having interest coverage in the bottom quartile of an industry, and the results are similar.

In Table 1.5, I investigate the relationship between human capital intensity and the evolution of capital structure and debt structure during the ten years prior to bankruptcy in OLS regressions. The specifications include event year, an indicator for the year relative to bankruptcy filing ranging from -10 to -1, capturing trends for all firms as they approach bankruptcy. In addition, they include *college share* and the interaction of *college share* and *event year*, which captures differential trends for human-capital-intensive firms.

Specifications (1) and (2) present the relationship between college share and the leverage. In both specifications, with the interaction of *college share* and *event year* being statistically significant at the 10% and 1% level, indicating that higher college share is associated with faster increases in firm leverage. A one-standard-deviation increase in college share is associated with a 0.003 to 0.004 per year faster increase in either unadjusted or industry-adjusted leverage.

The remaining specifications examine whether these increases in leverage are accompanied by changes in debt structure. Specifications (3) and (4) show that the maturity of debt shortens as firms approach bankruptcy by 0.02 per year. However, human-capital-intensive firms experience a slower decline in maturity by 0.002 to 0.003 per year for a one-standard-deviation increase in college share. As shown in (5) and (6), firms in general exhibit a slight increase in percentage of secured debt in the capital structure, but this effect is negated for by a one standard-deviation increase in college share.

Although the previous results suggest the possibility that human-capital-intensive firms are issuing more debt as they approach bankruptcy, increases in leverage could be driven either by debt issuance or by declines in assets. Thus, I next explore the sources of the leverage increase for human-capital-intensive firms. Table 1.5 investigates how human capital relates to changes in firm scale during the ten years prior to bankruptcy. The dependent variables for the first regression in each pair are year-to-year changes in log levels of debt, non-debt liabilities, asset size, capital expenditures, and employment. The second regression of each pair adjusts the change in log levels by subtracting its industry median among firms in the same three-digit NAICS industry in

[26]In the overall Compustat universe during the sample period, interest coverage has a correlation coefficient of -0.02 with respect to college share which is significant at the 1% level.

Compustat. Each specification also includes dummy variables for the normalized year relative to bankruptcy filing. The results show that college share is consistently associated with greater growth in total debt, non-debt liabilities, total assets, capital expenditures, and employment. The coefficients on college share are significant at the 1% level in all specifications. In particular, the correlations persist when the dependent variables are adjusted by industry, so the result does not appear to be driven by differences in the rate of firm growth across industries. Indeed, in each specification the effect of college share becomes even larger after adjusting for industry. Thus, although firms headed toward bankruptcy tend grow slower than other firms in the same industry, distressed firms which are human-capital-intensive exhibit smaller reductions in growth relative to their industries. While the magnitudes of the specifications I present are somewhat difficult to interpret in economic terms, the main result I highlight is that human-capital-intensive firms are more aggressive in continuing their borrowing and growth as they approach bankruptcy. I reach the same conclusion with similar significance levels when using logit specifications for increases in levels of each variable, percent changes in levels, or changes in levels normalized by the book value of assets in a base year.

It is important to note throughout the analysis that the results are subject to measurement error and omitted variable biases. Nonetheless, as I detail in Section 1.8, the results for the timing of bankruptcy and changes in firm scale are generally robust to the inclusion of additional controls for median industry measures of asset tangibility, Q, and R&D expenditures, as well as the inclusion of two-digit industry fixed-effects. Since pre-bankruptcy behaviors are likely to reflect the anticipated outcomes of bankruptcy itself, to obtain further insights into the role of human capital I next turn to the ultimate fates of firms at the resolution of bankruptcy.

1.5 Firm outcomes at the resolution of bankruptcy

In this section, I examine the relationship between human capital and the choice between emergence, acquisition, and liquidation of the firm at the resolution of bankruptcy. I first discuss univariate comparisons between the bankruptcy outcomes for firms with different degrees of human-capital-intensity. When splitting the sample by median college share, I find that low-college-share firms have a 55% likelihood of emerging from bankruptcy, a 35% chance of liquidating, and a 10% chance of being acquired in bankruptcy. In contrast, high-college-share firms have a 44% chance of emerging from bankruptcy, a 44% chance of liquidation, and a 12% chance of being acquired. Based on this initial comparison, human-capital-intensive firms are less likely to emerge from bankruptcy and are more likely to be liquidated, although their probability of acquisition is similar to that of low-human-capital firms. However, as uncovered in previous studies, other firm characteristics are also correlated with firm outcomes in bankruptcy, so I next turn to multivariate analyses.

In Table 1.8, I show that the relationship between college share and higher liquidation rates at the resolution of bankruptcy remains consistent in multivariate specifications. In these specifications, I conduct multinomial logit regressions where the dependent variable indicates the choice between emergence, acquisition and liquidation, with emergence as the omitted category. Specification (1) includes only college share as the independent variable in addition to a constant term, and the estimated coefficient indicates that the marginal effect of a one- standard-deviation increase is a five percent increase in the likelihood of liquidation. However, a similar increase in college share is associated with only a one percent increase in the probability of acquisition, with an insignificant coefficient.

Specification (2) includes firm characteristics which have been shown previously to impact bankruptcy decision-making. In particular, Denis and Rodgers [2007] find that larger firms and those with higher pre-bankruptcy leverage are more likely to emerge. This evidence is corroborated by Lemmon et al. [2009], who in addition find that industry-adjusted profitability is positively related to emergence. Following this previous literature, I include controls for log sales, leverage, and industry-adjusted profitability measured in the last fiscal year prior to the bankruptcy filing.[27]

Specification (2) shows that including the controls in addition college share yields similar results to previous studies. Both firm size and leverage are negatively and significantly correlated with liquidation probability. A one-standard-deviation increase in log sales is associated with a three percent decline in liquidation probability, while an analogous increase in leverage is associated with a nine percent decrease in liquidation probability. I also find that industry-adjusted profitability is negatively and significantly related to liquidation rates, with a one- standard-deviation increase in profitability being associated with a five percent decline in the probability of liquidation. Thus, although the controls may be imperfect proxies for factors unrelated to human capital which influence bankruptcy outcomes, the consistency of my findings indicate that the effect of college share does not seem to be driven by variation in firm size or the extent of financial or economic distress prior to bankruptcy.

In specification (3), the change in leverage between two and three years prior to bankruptcy and its interaction with college share are also included in the regressions. While the coefficients on the controls remain consistent with specification (2), the magnitude of the coefficient on college share is reduced by one third and remains insignificant.

Furthermore, the inclusion of human capital in regressions of bankruptcy outcomes has little impact on the results documented in previous literature. The lower liquidation rate for larger firms could be because of their greater scope for asset restructuring and operating improvements

[27]Using unadjusted profitability instead of industry-adjusted profitability yields similar results.

by divesting assets and refocusing their strategy.[28] Furthermore, it may also be more costly to liquidate larger firms due to the financing constraints of potential buyers (Shleifer and Vishny [1992], Acharya et al. [2007]). Firms with higher leverage may emerge at a greater rate both because they are more likely to suffer from financial as opposed to economic distress (Lemmon et al. [2009]), and because creditors have a greater incentive to promote continuation when insolvency is more severe and their claims become risky. Although Denis and Rodgers [2007] do not find a relationship between pre-bankruptcy profitability and liquidation probability, my results are consistent with those of Lemmon et al. [2009] and the intuition that pre-bankruptcy operating performance should be positively correlated with going concern value.

I now describe a basic framework for interpreting the results. In a simplified view of the liquidation decision in bankruptcy, firms are liquidated if their liquidation value exceeds the continuation value of their assets. If the control variables account for the continuation value of the firm prior to bankruptcy, then the relationship between human capital and liquidation rates which I observe could be due to either changes in continuation value during bankruptcy or to differences in liquidation value. As I mentioned in the introduction, I highlight two potential interpretations for the results.

First, I venture that human capital flight may lead to decline in the continuation value of the firm during bankruptcy. Analogous to physical assets which are collateral for creditors, a firm's human capital acts in a way as both an asset and a liability. In return for their service, the firm implicitly owes workers a stream of future wages. However, while bankruptcy law contains provisions such as the automatic stay[29] to prevent creditors from seizing collateral which might be critical to the firm's going concern value, the law does little to protect the going concern of human capital. As the firm's survival becomes uncertain due to financial distress, it may become optimal for workers to leave the firm for better employment opportunities, potentially destroying going concern value. Indeed, firms routinely claim that employee turnover is a significant threat that would hamper efforts to reorganize or sell the firm. It is important to note that even if human capital flight does influence the liquidation decision within bankruptcy, many other factors are also at work in the bankruptcy process. Indeed, if the pro-debtor bias of the U. S. bankruptcy code tilts toward over-continuation as suggested by Hotchkiss [1995] and Bradley and Rosenzweig [1992], then human capital flight may actually result in more efficient asset allocation.

The human capital flight interpretation can also help to shed light on the pre-bankruptcy behavior of human-capital-intensive firms described in the previous section. Due to the restrictions atten-

[28]As shown by Kalay et al. [2007], firms which emerge from Chapter 11 show significant reductions in size and improvements in profit margins, and Denis and Rodgers [2007] shows that reductions in asset size are correlated with the probability of emergence.

[29]The automatic stay is an injunction which prevents creditors from collecting on their debts or seizing assets without court permission.

dant to court supervision, it is likely that human capital flight is more important within bankruptcy than prior to bankruptcy.[30] If human capital flight is most salient within bankruptcy, then this interpretation may also help explain the aggressive growth of high-college-share firms prior to bankruptcy. If managers are aware that human capital flight is likely to trigger liquidation in the event of bankruptcy, they may optimally continue to invest and grow the firm during financial distress in hopes that the firm would recover.

The second interpretation of the results is that human-capital-intensive firms have assets which are more readily redeployable, thereby resulting in higher liquidation values. As illustrated in Table 1.3, industries with the highest levels of human capital generally center around high-skilled services, so human-capital-intensive firms may tend to have generic physical assets such as real estate and computer equipment which are more easily repurposed[31]. Consistent with this idea, Baird and Rasmussen [2002] argue that human-capital-intensive firms have few dedicated assets, so it is less worthwhile to reorganize them as going concerns during bankruptcy.

I next examine the post-bankruptcy performance of firms which emerge from Chapter 11. By comparing the relationships between human capital and outcomes within bankruptcy versus post-bankruptcy outcomes, I can distinguish between two different ways in which human capital impacts firm outcomes. If human-capital-intensive firms which enter bankruptcy are simply less viable overall compared with low-human-capital firms, then human capital should be correlated with higher post-bankruptcy refiling and liquidation rates and lower rates of continuation. However, if the selection process within bankruptcy is different for human-capital-intensive firms, then they could enjoy higher levels of continuation after emerging from bankruptcy despite their higher propensity for liquidation within bankruptcy.

1.6 Post-bankruptcy performance and survival for emerging firms

The post-bankruptcy survival of firms which emerge from bankruptcy can act as a test of whether the bankruptcy process results in over-continuation of firms which should be liquidated within

[30] Bankruptcy does have several specific implications for employees. Wages and benefits garnered prior to bankruptcy filing receive priority status up to $4,925, which are paid in full at the end of bankruptcy proceedings. As part of the first day of bankruptcy court proceedings, many firms obtain a court order to continue compensating their employees under the arrangements which were in place prior to bankruptcy. However, the increased scrutiny of the court is likely to limit firms' abilities to retain employees by raising wages, increasing benefits and perquisites, or providing retention incentives. In particular, when creditors have security interest in a firm's working capital, the debtor-in-possession must obtain permission from the court and agreement from the creditors to use proceeds of this "cash collateral" to fund ongoing operations including employee wages.

[31] As argued by Autor et al. [2003], the increased demand for educated workers may be driven by strong complementarities between computers and the nonroutine cognitive tasks prevalent in high-skilled services.

Chapter 11. As documented by Hotchkiss [1995], a large fraction of firms which emerge from bankruptcy continue to perform worse than their industry peers, and many are liquidated or re-enter bankruptcy within five years of emergence. These results have been interpreted to suggest that pro-debtor biases in the bankruptcy process lead to the over-continuation of firms which should be shut down. Based on the results from the previous section, the extent of asset reallocation in bankruptcy may differ by human capital, but the specific channels for human capital's impact have different implications for post-bankruptcy outcomes.

If the higher liquidation rates for human-capital-intensive firms result only from the dissipation of going concern value due to human capital flight, then these firms would not be expected to fare better after bankruptcy compared with low-human-capital firms. Alternatively, if the higher liquidation rates documented in the previous section arise from greater asset redeployability, then human-capital-intensive firms may fare better conditional on emerging from bankruptcy due to the more stringent selection process within bankruptcy. To adjudicate between these possible relationships, I analyze post-bankruptcy performance and survival for the subset of 441 filings in which firms emerge from Chapter 11 as public companies, and I will present results for both profitability during the five years following emergence from bankruptcy and on post-bankruptcy firm survival.

Before I discuss my results for post-bankruptcy performance, an important note for comparison is that the operating performance for my sample of emerging firms is substantially better both before and after bankruptcy compared with those studied in Hotchkiss [1995]. While Hotchkiss [1995] finds that between 36 and 40 percent of emerging firms continue to experience negative operating income in the five years after emergence, this measure is between 19 percent and 24 percent in my sample.[32]

First, I consider whether human-capital-intensive firms improve their profitability at a faster rate after emerging from bankruptcy. In unreported results, I find that college share is associated with faster increases in profitability in the *emerged public* sample during the five after emerging from bankruptcy, although this relationship is insignificant. One issue with this analysis is that profitability can only be observed for surviving firms, while a major marker of poor performance is that firms drop out of the sample due to liquidation or refiling after initial emergence from bankruptcy. For more definitive evidence, I examine the outcomes among firms in the *emerged public* sample as of five years after emergence from bankruptcy. Of the 441 firms in my *emerged public* sample, 74 (17%) refile for Chapter 11 within five years of emergence, while 68 (15%)

[32]These differences are somewhat mitigated when I restrict my sample either to firms filing before 1989 (consistent with Hotchkiss's sample period) or restricting to firms with total assets less than $400 million. Within these subsamples, about 30 percent of firms continue to experience negative operating income in the five years after emergence. The remaining discrepancy is likely to result from sample selection; Hotchkiss [1995] supplements Compustat data with that collected from SEC filings, and it is likely that firms with missing data from Compustat have relatively lower operating performance.

are acquired, and 4 (1%) are liquidated, and 26 (6%) have unknown outcomes. The overall rates of post-bankruptcy distress are somewhat lower in my sample than those studied previously. By comparison, Hotchkiss [1995] finds that 25 percent of firms re-enter distress through a private restructuring, bankruptcy, or liquidation within five years of emergence. Lemmon et al. [2009] find that 32 percent of firms in their sample are acquired or re-file for bankruptcy within three years of emergence. While I focus on firm outcomes at the end of five years after initial emergence from bankruptcy, I obtain qualitatively similar results when varying the horizon after emergence between 3 and 7 years, and the results become stronger as the horizon increases.

The results corroborate the evidence from firm profitability discussed above that human-capital-intensive firms exhibit greater degrees of post-bankruptcy success. On a univariate basis, human-capital-intensive firms exhibit a greater probability of survival upon emergence from bankruptcy and a lower probability of refiling within five years. When splitting the *emerged public* sample by median college share, I find that high-college-share firms have a 67% likelihood of continuation, a 2% likelihood of liquidation, a 12% likelihood of refiling, and a 17% chance of being acquired as of five years after emerging from bankruptcy. In contrast, low-college-share firms have a 58% likelihood of continuation, a 1% likelihood of liquidation, a 26% likelihood of refiling, and a 16% chance of being acquired.

To analyze the relationship between post-bankruptcy outcomes and human capital on a multivariate basis, Table 1.9 presents results from multinomial logit regressions in which the dependent variable indicates the choice between the three possible post-bankruptcy outcomes, with continuation being the omitted category. Because very few firms are liquidated out of court subsequent to emergence, these two outcomes are collapsed into one category, which I call "failure". In specification (1), I confirm the univariate results by including only college share as a dependent variable in addition to the constant term. The results indicate that a one-standard-deviation increase in college share is associated with an eight percent reduction in the probability of post-bankruptcy failure.

To account for other variables which may be correlated with post-bankruptcy outcomes, specifications (2) and (3) include controls for log sales, leverage, and profitability measured in the fiscal year prior to bankruptcy filing.[33] In specification (2), the controls are added without the college share measure to establish a basis for comparison. Although the focus of my analysis is on the role of human capital, it is worth noting the patterns of results for outcomes within bankruptcy and post-bankruptcy outcomes for the control variables. If the liquidation decision within bankruptcy is made according to going concern value, the factors which are associated with liquidation within bankruptcy should also predict poor post-bankruptcy outcomes. Interestingly, the correlation between firm size and post-bankruptcy success has the opposite sign as for survival within bankruptcy

[33]In unreported regressions, the results are similar when measuring controls in the first fiscal year after emergence from bankruptcy instead of prior.

(see Table 1.8), and its magnitude is economically significant. Evaluated at the means of the other variables, a one- standard-deviation increase in log sales is associated with an eight percent increase in the probability of failing within five years. Thus, the results may indicate that frictions in the bankruptcy process may favor the over-continuation of larger firms. Although leverage is also positively correlated with post-bankruptcy failure in contrast with its negative relationship with liquidation within bankruptcy, the coefficient is insignificant both economically and statistically. However, consistent with the idea that pre-bankruptcy profitability proxies for going concern value, this variable is negatively related both to liquidation within bankruptcy and post-bankruptcy failure. A one-standard-deviation increase in industry-adjusted profitability is associated with an eight percent decrease in the likelihood of post-bankruptcy failure.

Specification (3) shows that when both college share and the control variables are included in the regression, the estimated effect of college share is slightly reduced but remains both economically and statistically significant. In this specification, a one-standard-deviation increase in college share is associated with a seven percent decline in the probability of post-bankruptcy failure. The marginal effects of the control variables also remain very similar to their levels from specification (2). Overall, the stability of the estimates in the three specifications indicates that the relationship between human capital and post-bankruptcy outcomes does not appear to be driven by its correlation with firm size, pre-bankruptcy leverage, or pre-bankruptcy profitability. Nonetheless, the results may still result from omitted variable bias, and I explore additional robustness checks in Section 1.8.

I now revisit my two main interpretations with respect to the observed relationship between college share and post-bankruptcy outcomes. Because higher levels of college share are associated with higher levels of liquidation at the resolution of bankruptcy yet lower failure rates upon emergence, the selection process within bankruptcy is likely to differ for these firms.[34] As discussed above, human capital flight may be one reason that human-capital-intensive are liquidated in bankruptcy. However, if employees depart from firms indiscriminately, then I should not observe lower post-bankruptcy failure rates. A more likely scenario is that employees choose to stay with firms which have higher continuation value, while hastening the liquidation of less viable firms. Hence, human capital flight may play a role not only in the overall liquidation rates of firms in bankruptcy, but in the differential selection of which firms emerge. Another reason that the selection process in bankruptcy may differ for human-capital-intensive firms is that their assets are more easily redeployed. Under this interpretation, only human-capital-intensive firms with relatively high continuation values would emerge from bankruptcy because the value of their assets under alternative uses is high. An important implication of both of these selection mechanisms is

[34]Otherwise, the higher liquidation rates would be indicate lower unobserved firm quality, leading to higher failure rates upon emergence.

that the inefficient continuation of firms in the bankruptcy process appears to be less salient for human-capital-intensive firms.

1.7 Extension: general and firm-specific human capital

In this section, I re-examine my main results with the inclusion of employer tenure as a proxy for firm-specific human capital. The construction and justification of employer tenure is described in Section 1.3. When both college share and employer tenure are included in the specifications, I interpret college share as a proxy for general human capital, while employer tenure is likely to be more correlated with firm-specific human capital. However, it is important to note that employer tenure could be a proxy for omitted variables in addition to firm-specific human capital, so these results should be interpreted with caution.

The impact of general versus firm-specific human capital is likely to differ with respect to my two main interpretations for the relationship between human capital and the bankruptcy process. For example, it could be that human capital flight is less likely when employees possess greater levels of firm-specific as opposed to general human capital, since the value of their human capital is worth less to outside employers. Another possibility is that firm-specific human capital reduces the redeployability of a firm's assets through cospecialization between human and non-human assets. Although these are two plausible interactions, the distinct effects of general versus firm-specific human capital remain empirical concerns. Thus, the following analysis can help shed additional light on the underlying mechanisms through with which human capital interacts with the bankruptcy process.

Overall, the analysis in this section indicates that firm-specific human capital seems to play a distinct role in the bankruptcy process compared with that of general human capital. I now evaluate the patterns in the bankruptcy process associated with employer tenure in light of the two interpretative frameworks I have discussed throughout this study: human capital flight and the relationship between human capital and asset redeployability.

First, I discuss the implications of my results with respect to human capital flight. Ex ante, it is unclear whether human capital flight should be a more important problem for firms with firm-specific or general human capital. While likelihood of human capital flight may be higher if employees have general human capital,[35] the value lost may be greater for firms with firm-specific human capital. My results show that while firms with higher levels of firm-specific human capital employ general cutbacks in firm scale, they seem relatively more reluctant to scale back employment than assets, borrowing, or capital expenditures. Furthermore, because employment reductions are coincident with general cutbacks, it is likely that separations generally occur because of

[35] See Hall and Lazear [1984] and Hashimoto [1981]

layoffs initiated by firms rather than voluntary flight by employees. The insignificant and negative coefficient on employer tenure with respect to liquidation within bankruptcy is also consistent with the idea that human capital flight is less worrisome when human capital is firm-specific.

Next, I discuss how firm-specific human capital might relate to asset redeployability. Ceteris paribus, firms with higher levels of firm-specific human capital may have lower asset redeployability due to cospecialization between human and nonhuman assets. Firms with high levels of firm-specific human capital tend to involve sophisticated manufacturing processes (see Section 1.3), so these assets may be worth little to a buyer who intends to use them for another purpose. Another factor which impacts asset redeployability is that it may be more costly for firms with high firm-specific human capital to dismantle their workforce following asset sales.[36]

At first, the greater willingness of firms with firm-specific human capital to scale back their assets prior to bankruptcy seems inconsistent with the idea that their assets are less easily redeployed. However, a controlled reduction in firm size may be necessary to forestall bankruptcy (with its attendant risks of complete liquidation) in hopes that most of the firm would remain intact. Firms with high levels of firm-specific human capital may thus be strategically selling off non-core assets in order to maintain the viability of the firm as a whole. As argued by Shleifer and Vishny [1992], industry distress is particularly worrisome when assets have low redeployability since potential buyers are likely to be financially constrained, so the faster entry into bankruptcy associated with firm-specific human capital after industry distress may be related to the protection that bankruptcy can afford against asset fire-sales. Indeed, the negative (albeit insignificant) relationship between employer tenure and liquidation within bankruptcy is tentative evidence firm-specific human capital is associated with lower redeployability. In a related paper, Alderson and Betker [1995] find that firms with higher liquidation costs emerge from Chapter 11 with capital structures that make re-entry into financial distress less likely. Under the hypothesis the firm-specific human capital is correlated with higher liquidation costs, my result that employer tenure is related to lower levels of post-bankruptcy failure is consistent with this previous finding.

The tentative results and discussion in this section provide some initial evidence that the distinction between general and firm-specific human capital makes a difference in the bankruptcy process. With respect to my two main interpretations, human capital flight seems to be a more important channel with respect to general human capital, while the relationship between human capital and asset redeployability seems to drive the results when firm-specific human capital is important.

[36]In order to incentivize workers to make the necessary investments in firm-specific human capital and to subsequently retain them, firms must promise sufficient long-term rewards which may include provisions for higher severance pay and benefits upon liquidation. Although these contracts are sometimes reneged upon or renegotiated during bankruptcy, employee claims are often a significant part of a firm's liabilities and are likely to be higher when firm-specific investments are important even after renegotiation.

1.8 Robustness checks

As I discussed in Section 1.3, the *college share* measure is likely to be correlated with differences in human-capital-intensity across industries. However, industries which differ by human capital are also likely to differ in other ways, leaving room for the results I've presented to be driven by omitted variables. In particular, firms with high college share tend to have more intangible assets, higher growth opportunities, and more R&D expenditures. In this Section, I explore how my main results are affected by considering the impact of potential omitted variables.

As proxies for intangibles and growth opportunities, I include median industry levels of tangibility (defined as PPE/book value), Q, and R&D expenditures / book value. For ease of exposition, I will refer to $intangibility$ in the following discussion and in the tables as the negative of tangibility. Because the value of assets and expenditures fluctuate considerably for firms approaching bankruptcy, industry-level measures are likely to be better proxies for the underlying importance of intangibles and growth opportunities than firm-level measures. However, in unreported results, my results are also robust to the inclusion of firm-level measures of intangibility, Q, and R&D.

In Table 1.11, I document the relationship between college share and firm outcomes in bankruptcy by including the additional controls in specification (3) of Table 1.8. Although it becomes insignificant or only marginally significant with the inclusion of the controls, the coefficient on college share remains positive in all specifications with a magnitude comparable to that without the inclusion of controls. Similar to college share, industry Q , intangibility, and $R\&D$ expenditures are all positively correlated with the probability of liquidation within bankruptcy. A one-standard-deviation increase in industry Q is associated with a three percent increase in liquidation rates, a one- standard-deviation increase in industry intangibility is associated with a five percent increase in liquidation rates, and a one-standard-deviation increase in industry R&D expenditures is associated with a less than one percent increase in liquidation rates. The results suggest that human capital has additional explanatory power beyond that of intangible assets and growth opportunities.

Another technique for accounting for omitted variables which could be correlated with human capital is through two-digit industry fixed-effects. As described in Section 1.3, college share is constructed at the three-digit NAICS industry level, and industries which vary by college share are also different on a variety of other dimensions. Furthermore, the inclusion of proxies for omitted variables does not capture variation in *unobservable* dimensions, which can be accounted for by estimating industry fixed-effects. While college share varies at the level of the 92 three-digit NAICS industries, the two-digit NAICS groups industries into broader classifications. For instance, within the health care and social assistance industry (NAICS 62), nursing and residential care facilities (NAICS 623) have a college share of fifteen percent, while social assistance (NAICS 624) has a college share of thirty seven percent. A major drawback to this strategy is that the two-digit industries absorb much of the variation in human capital, so inclusion of industry fixed-effects significantly reduces the power of these tests.

Table 1.12 shows results for bankruptcy outcomes and post-bankruptcy outcomes with the inclusion of two-digit NAICS fixed-effects, replicating the specifications in column (3) of Tables 1.8 and 1.9.[37] Compared with the result from Table 1.8, the coefficient on college share in specification (1) with respect to liquidation within bankruptcy remains positive, but it becomes insignificant with the inclusion of fixed-effects and the marginal effect is reduced by 34 percent. The marginal effect of college share in the post-bankruptcy regression in specification (2) is reduced in magnitude by 79 percent compared to the result in Table 1.9, but it remains significant at the one percent level. Thus, while the magnitudes and significance of the effects for college share in the regressions are reduced, my main conclusions remain consistent in specifications that include two-digit industry fixed-effects.

1.9 Conclusion

This study presents an analysis of the role of human capital in the process of bankruptcy for a large sample of public companies that filed for Chapter 11 between 1980 and 2003. Using a measure of human capital intensity constructed from the Current Population Survey, I document several patterns in firm outcomes and behavior during the lead-up to bankruptcy, at the resolution of bankruptcy, and upon emergence from bankruptcy. Human-capital-intensive firms persist for a longer period with low interest coverage before filing for bankruptcy, but file faster following low stock returns and industry distress. Human-capital-intensive firms also exhibit more aggressive growth in borrowing, assets, investment, and employment compared with low-human-capital firms as they approach bankruptcy. They experience higher liquidation rates within bankruptcy and lower failure rates upon emergence. I show that these results are robust to a variety of different specifications as well as controls for several potential omitted variables and industry fixed-effects.

I propose two potential interpretations for the patterns I document. First, human capital flight may dissipate the going concern value of human-capital-intensive firms in bankruptcy, resulting in their higher liquidation rates. Moreover, their lower post-bankruptcy failure rates are consistent with the possibility that human capital flight may lead to the selective liquidation of firms that are less economically viable. Second, it could be that the assets of human-capital-intensive firms are more readily redeployed, so they experience fewer liquidation costs and thus higher rates of liquidation within bankruptcy. Greater asset redeployability could also facilitate the liquidation of less-viable firms within bankruptcy. Both of these interpretations suggest that the growing importance of human capital may have implications for the effectiveness of Chapter 11 in asset

[37] Results are qualitatively similar for the other specifications in Tables 1.8 and 1.9

allocation in the economy. While I present preliminary evidence that both channels may be at work, further research is necessary to more thoroughly examine the role of human capital in corporate bankruptcy.

1.10 Data Appendix

1.10.1 Current Population Survey

I construct human capital measures using the Current Population Survey (CPS), which is a monthly survey conducted by the Bureau of the Census. To obtain a universe of workers relevant for Compustat firms, I restrict the CPS sample to workers in the private sector who are not self-employed. Because the CPS does not provide information on employer identities, I use industry classifications to construct proxies which can be matched to Compustat. In the CPS, the observations from 1981-1982 used the 1970 census industrial classification; those from 1983-1991 used the 1980 census industrial classification; those from 1992-2002 used the 1990 census industrial classification; and those from 2003-2004 used the 2000 census industrial classification. The 2000 census industrial classification is matched to the 2002 NAICS classification through crosswalks provided by the census bureau, and NAICS classifications are obtained for earlier census industries by using crosswalks to the 2000 classification. Through this process, each census industry is matched to a NAICS industry at the two to six-digit level. Using the resulting industry matches, I construct the *college share* and *employer tenure* measures for each two- and three-digit NAICS industry.[38]

1.10.2 College share

For the *college share* measure, I use the annual March supplement which has been extensively studied in the labor literature and includes about 100,000 individuals in each year, covering data on employment characteristics such as the earnings, industry, and occupation of each worker, as well as demographic characteristics such as age, sex, and ethnicity, and educational attainment. I obtain a consistent data series for the March supplement from the Integrated Public Use Microdata Series (IPUMS). I classify workers as college-educated if they have at least sixteen years of educational attainment. In each year t for each individual j, individual-level survey weights $w_{j,t}$ are provided in the CPS which represent the inverse probability of selection into the sample from the general population. Using these weights and the industry matching procedure described above, the college share measure for each three-digit NAICS industry i and year t is defined as

[38]Although the measures used in the analysis are at the three-digit level, two-digit measures are substituted when the census classification is matched at the two-digit NAICS level.

$$CollegeShare_{i,t} = \frac{\sum_{j \in i} w_{j,t} college_{j,t}}{\sum_{j \in i} w_{j,t}}$$

where $college_{j,t}$ is a dummy variable for college education for each worker. Each March survey describes individual characteristics for the previous calendar year, so I match each year t in Compustat to the CPS survey in year $t + 1$. Thus, I used the surveys from 1981 to 2004 to construct the *college share* measure used in the analysis.

1.10.3 Employer tenure

Data on employer tenure are reported in the CPS mobility supplements from January 1981, 1983, 1987, 1991, 1996, 1998, 2000, and 2002 and the contingent and alternative employment arrangement supplements (CAEAS) in February 1995 and 1997. I obtain data from these supplements from UNICON and use the same year and industry conventions as for the *college share* measure described in the previous section. These supplements have been used in a substantial prior literature on employee tenure (Farber [1999], Jaeger and Stevens [1999], Farber [2007]).

The *employer tenure* variable measures the length of time an employee has worked for her current employer at the time of the survey, so it is important to note that this measure underestimates the length of completed employment spells. The survey question has also changed over time. In 1981, the survey question asked when the individual started working at her present job. From 1983 to 1991 1996, and 1998 to 2004, the survey question asked for the length of time the individual has worked continuously for her present employer. In 1995 and 1997, the survey question asked how long the individual has worked for her employer, without mention of the continuity of employment. Among these changes, the most substantial break occurred between the 1981 and subsequent surveys in the shifts from "job" to "employer" and from the date started to the length of time completed. However Jaeger and Stevens [1999] document that the tenure series remain stable across the two surveys as a result of the offsetting effects of the two questions.

I define the *employer tenure* measure for each three-digit NAICS industry i and year t as

$$EmployerTenure_{i,t} = \frac{\sum_{j \in i} w_{j,t} TenureVar_{j,t}}{\sum_{j \in i} w_{j,t}}$$

where $TenureVar_{j,t}$ is the response to the survey question described above for year t. For tenure measures taken from the displaced worker supplements in 1996, 1998, 2000, and 2002, $w_{j,t}$ is the displaced worker supplement weight. In other years, $w_{j,t}$ is the basic supplement weight.

Table 1.1
Pre-bankruptcy Firm Characteristics by Chapter 11 Outcome

The table presents firm characteristics in the last fiscal year prior to filing for Chapter 11 for 1493 filings between 1980 and 2003. The sample is divided into five outcomes at the resolution of Chapter 11. *Emerged public* indicates that an operating company emerged from Chapter 11 and was subsequently listed in the Compustat data set. *Emerged private* indicates that an operating company emerged from Chapter 11, but was not subsequently listed in Compustat. *Liquidated* indicates that the firm was sold piecemeal in a Chapter 11 liquidation or a conversion to Chapter 7. *Acquired* indicates that the majority of the firm's assets were acquired as a going concern by a single buyer. Firm outcomes are determined from the WebBRD, Bankruptcydata.com, court documents, and news searches. *Unknown* indicates firms whose outcomes could not be determined from these sources. ***,**,* denote that the difference in mean (median) characteristics between the outcome indicated and the sample of emerged firms (both public and private) is significant at the 0.01, 0.05, and 0.10 levels based on a t-test (Wilcoxon rank sum test).

	(1) Emerged public	(2) Emerged private	(3) Liquidated	(4) Acquired	(5) Unknown	(6) Full sample
Panel A: Number and proportion of filings by outcome						
Number of filings	441	267	566	159	60	1493
% of filings	30%	18%	38%	11%	4%	100%
Panel B: Mean (median) firm characteristics in year prior to Chapter 11 filing						
Sales ($MM)	926.34	953.06	359.17***	192.06**	491.79	613.92
	(289.87)	(127.66)	(94.52)***	(72.14)***	(35.99)***	(125.63)
Profitability	−0.23	−0.19	−0.18	−0.65	−0.16	−0.25
	(0.04)	(0.01)	(−0.03)***	(−0.03)***	(−0.06)***	(0.001)
Total debt / assets	0.94	0.75	0.54***	0.76	0.55	0.71
	(0.66)	(0.67)	(0.48)***	(0.53)**	(0.43)***	(0.55)
Total liabilities/ assets	1.53	1.17	0.93***	1.36	1.08	1.19
	(0.96)	(0.97)	(0.82)***	(0.86)***	(0.85)**	(0.88)
Interest coverage	−8.60	−7.44	−5.26	−10.54	−14.30	−7.56
	(0.57)	(0.24)	(−0.52)***	(−0.23)***	(−0.98)***	(0.04)
Panel C: Duration of Chapter 11						
Months in Chapter 11	16.87	17.01	21.74***	17.73	29.19***	19.19
	(13.00)	(13.93)	(15.98)***	(13.53)	(20.80)***	(14.33)

Table 1.2
Human Capital and the Outcomes of Financial Distress

Panel A: Firm outcomes 10 years after entering financial distress.
Firm outcomes are defined by Compustat exit codes for 19,053 firms that enter financial distress between 1980 and 2003. The onset of financial distress is defined as the first of two years in which interest coverage falls below 1, or any year in which interest coverage falls below 0.8. The sample is split by median college share (21%) into high- and low-college-share groups.

10-year outcome	Low college share N=9,029 %	High college share N=10,024 %	Total %	N
Merger	15.6	20.4	18.1	3,457
Bankruptcy	11.2	7.2	9.1	1,726
Liquidation	1.8	1.4	1.6	303
Buyout	2.4	1.2	1.8	341
Other	29.0	26.1	27.5	5,233
Survived	40.0	43.7	42.0	7,993
Total	100.0	100.0	100.0	19,053

Pearson chi2(5) = 221.46 ; Pr = 0.00

Panel B: Years between onset of distress and bankruptcy.
Sample of 1484 firms is split by median college share (20%). Statistics are presented for the length of time in years between the beginning of the fiscal year for the first year of financial distress as measured by either low interest coverage as defined in Asquith et al. [1994], or as having interest coverage below the contemporaneous 25th percentile in a three-digit NAICS industry.

	Low college share N=728		High college share N=756	
	Mean	Median	Mean	Median
Low int cov	3.72 * **	2.99 * **	4.41 * **	4.08 * **
Int cov < p25 ind	4.32 * *	3.95 * **	4.77 * *	4.39 * **

* significant at 10%; ** significant at 5%; *** significant at 1%

Figure 1
Trends in Human Capital

Average college share and employer tenure among full-time workers in the private sector in the Current Population Survey (CPS). College share and employer tenure are defined in Section 1.3.

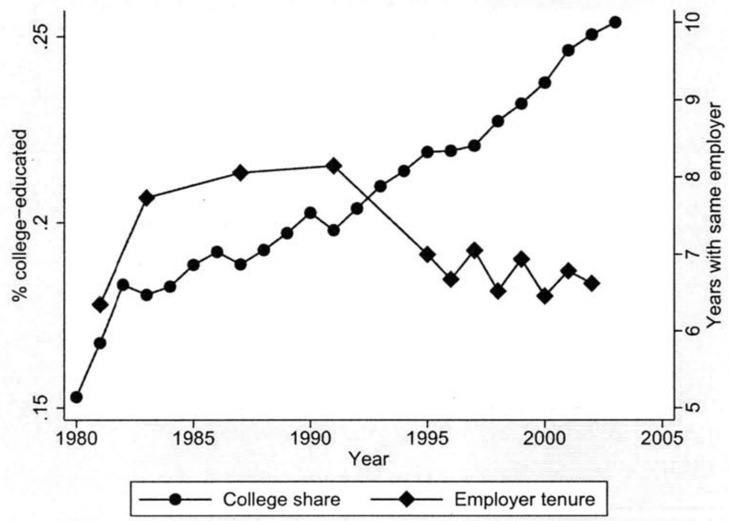


Table 1.3
Industry Composition and Human Capital Measures

The table presents the number and percentage of firms in each two-digit NAICS industry 1493 public firms which file for Chapter 11 between 1980 and 2003. The percentage of firms in each industry in the full Compustat sample between 1980 and 2003 is also presented. Average college share and employer tenure for each industry are as defined in Section 1.3.

Industry	# Firms Ch 11 sample	% Firms Ch 11 sample	% Firms Compustat sample	College share	Employer tenure
Accommodation and Food Services	64	4.3%	3.2%	7.6%	4.1
Administrative and Support and Waste Management and Remediation Services	39	2.6%	3.8%	19.5%	4.4
Agriculture, Forestry, Fishing and Hunting	5	0.3%	0.6%	6.8%	6.7
Arts, Entertainment, and Recreation	15	1.0%	1.4%	18.2%	5.0
Construction	36	2.4%	2.5%	8.6%	5.6
Educational Services	5	0.3%	0.5%	54.8%	7.0
Food, Beverage, Tobacco, Textile, Apparel, and Leather Manufacturing	99	6.6%	5.7%	9.9%	8.5
Health Care and Social Assistance	61	4.1%	2.9%	27.7%	5.9
Information	204	13.7%	16.4%	31.2%	7.8
Metal, Machinery, Electronics, Transportation, and Miscellaneous Manufacturing	354	23.7%	30.2%	19.2%	9.3
Mining	73	4.9%	7.8%	20.9%	9.0
Other Services (except Public Administration)	7	0.5%	0.8%	15.4%	5.7
Postal Service, Warehousing	5	0.3%	0.3%	9.2%	6.1
Professional, Scientific, and Technical Services	43	2.9%	7.7%	48.9%	5.3
Public Administration	1	0.1%	0.0%		
Real Estate and Rental and Leasing	39	2.6%	4.5%	22.7%	5.0
Retail trade in Motor Vehicles, Furnishings, Electronics, Building Materials, Food, Personal Care	118	7.9%	3.8%	11.1%	5.8
Retail trade, General and Miscellaneous	81	5.4%	3.1%	13.2%	5.8
Transportation and Warehousing	63	4.2%	3.4%	13.0%	8.5
Wholesale Trade	75	5.0%	6.2%	22.1%	7.0
Wood product, Petroleum, Chemical, Plastics, and Nonmetallic Mineral Product Manufacturing	106	7.1%	13.1%	18.7%	9.5
Total	1493	100.0%	100.0%	19.8%	7.5

Table 1.4
Correlations Between Human Capital and Firm Characteristics

The table presents correlations between college share, employer tenure, and firm characteristics averaged within the 70 three-digit NAICS industries represented in a sample of 1493 Chapter 11 filings by public firms between 1980 and 2003. Average college share and employer tenure for each industry are as defined in Section 1.3. Equally-weighted means for firms in each industry are calculated for *Q*, *tangibility*, *leverage*, *R&D/book*, and *profitability* measured two fiscal years before filing. *Q* is defined as (book value of assets + market value of common equity - book value of equity - book value of deferred taxes) / book value of assets. *tangibility* is defined as net property, plants and equipment (PPE) / book value of assets. *leverage* is defined total liabilities / total assets. *R&D/book* is R&D expenses / total assets. *profitability* is EBITDA / total assets.

Variables	College share	Employer tenure	Log sales	Q	Tang	Leverage	R&D/book	Prof
College share	1.000							
Employer tenure	−0.120	1.000						
Log sales	−0.342	0.243	1.000					
Q	0.217	−0.032	−0.413	1.000				
Tangibility	−0.078	0.138	0.109	−0.161	1.000			
R&D/book	0.268	−0.063	−0.616	0.686	−0.243	1.000		
Leverage	−0.031	0.013	0.012	0.311	0.028	−0.112	1.000	
Profitability	−0.114	0.115	0.372	−0.831	0.245	−0.561	−0.172	1.000

Table 1.5
OLS Regressions Relating Leverage and Debt Structure to Human Capital

The table presents estimates of correlations between *college share* and the leverage and debt structure of firms during the ten years prior to bankruptcy for 1493 public firms that file for bankruptcy between 1980 and 2003. See Section 1.3 for the definition of *college share*. Industry-adjusted measures are calculated by subtracting the contemporaneous median of the dependent variable among firms in the same three-digit NAICS industry in the Compustat database. *Event year* indicates the year relative to filing year, ranging from -10 to -1. Dependent variables are winsorized at the 5% level. T statistics are presented in brackets, and standard errors are clustered at the industry level.

	(1)	(2)	(3)	(4)	(5)	(6)
	Leverage		ST debt / total debt		Secured debt / total debt	
	Unadj	Ind-adj	Unadj	Ind-adj	Unadj	Ind-adj
Event year	0.023	0.019	0.020	0.021	0.0041	0.0076
	[8.14]***	[7.52]***	[5.42]***	[7.00]***	[0.89]	[1.70]*
College share	−0.17	0.20	0.21	−0.27	−0.32	0.055
	[−1.58]	[1.49]	[1.83]*	[−3.06]***	[−2.28]**	[0.45]
College share X event year	0.033	0.044	−0.023	−0.032	−0.055	−0.036
	[1.70]*	[2.86]***	[−1.26]	[−2.16]**	[−2.48]**	[−1.69]*
Constant	0.89	0.25	0.35	0.21	0.41	0.12
	[37.47]***	[10.15]***	[14.08]***	[11.00]***	[12.00]***	[3.64]***
Observations	10329	10329	9846	9846	8074	8074
R-square	0.084	0.070	0.027	0.019	0.0051	0.0052

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.6
OLS Regressions Relating Pre-bankruptcy Changes in Firm Scale to Human Capital

The table presents estimates of correlations between *college share* and pre-bankruptcy changes in log levels of total debt, non-debt liabilities (total liabilities minus the book values of short-term and long-term debt), the book value of assets, capital expenditures, and total employment for 1493 public firms which file for bankruptcy between 1980 and 2003. See Section 1.3 for the definition of *college share*. The dependent variable in each specification is the change in log levels of each measure from year to year for an individual firm. *Event year* denotes the year relative to bankruptcy filing, ranging from -9 to -1. T statistics are presented in brackets, and standard errors are clustered at the industry level.

	(1) Total debt	(2) Non-debt liabilities	(3) Book value	(4) Capital expenditures	(5) Employment
Event year	0.014	−0.000087	−0.024	−0.039	−0.016
	[3.43]***	[−0.04]	[−9.83]***	[−9.67]***	[−5.96]***
College share	0.68	0.59	0.57	0.66	0.27
	[2.88]**	[3.03]**	[2.66]**	[3.12]**	[1.96]
Constant	0.16	0.082	−0.056	−0.21	−0.059
	[4.34]***	[2.41]*	[−1.90]	[−5.55]***	[−2.13]*
Observations	8256	8777	8844	8432	8082
R-square	0.0062	0.011	0.017	0.012	0.0084

Table 1.7
Pre-bankruptcy Debt Issuance

Percentage of firms that issued each type of debt in each year relative to filing year, split by median college share. Sample consists of 405 firms in the bankruptcy dataset that are rated by Moody's.

	Maturity						Type							
	Subordinated		Senior unsecured		Senior secured		Bank		Bonds		Convertible		All debt	
Human capital :	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High
Year														
-10	4.8	6.7	4.0	6.7	0.8	1.1	0.8	0	5.6	5.6	1.6	5.6	6.5	12.2
-9	5.9	5.9	3.7	5.9	1.5	2.9	0	1.0	8.1	6.9	2.2	3.9	11.1	11.8
-8	9.9	8.0	5.9	8.8	2.0	2.7	0*	1.8*	13.2	13.3	3.3	3.5	15.8	17.7
-7	9.4	8.2	9.4	6.7	1.8	3.0	0.6	1.5	17.0	11.2	3.5	3.7	18.7	17.2
-6	9.0	9.0	7.4	9.0	5.3	4.5	5.3	2.6	11.7	16.1	3.7	4.5	19.1	20.6
-5	12.7**	6.5**	6.8	7.7	6.8	4.1	3.9	4.1	19.5**	10.7**	1.5	3.0	24.9**	16.0**
-4	13.7	13.4	7.5	12.3	6.6	3.9	6.6	3.4	16.0	19.6	3.8	5.6	23.1	25.7
-3	12.5	14.1	11.1**	19.5**	6.9	9.2	4.2	6.5	20.8**	30.3**	4.2	5.9	29.2*	37.8*
-2	9.3	13.4	11.6	14.0	8.8	5.0	7.0	3.9	18.6	24.0	1.4*	4.5*	26.0	28.5
-1	5.2	4	3.6**	9.3**	2.6	4.7	2.6	3.3	8.3	12	1.6	3.3	10.9	16
Total	9.6	9.4	7.5	10.6	4.8	4.5	3.5	3.2	14.6	16.3	2.7	4.4	19.6	21.7

Table 1.8

Multinomial Logit Regressions Relating Bankruptcy Outcomes to Human Capital

The table presents estimates of the relationship between *college share* and bankruptcy outcomes from multinomial logit regressions of firm outcomes at the resolution of Chapter 11 for firms which filed between 1980 and 2003. See Section 1.3 for definition of *college share*. The dependent variable indicates whether the firm emerges from bankruptcy as an independent entity, was liquidated within bankruptcy, or was acquired within bankruptcy, with emergence as the omitted category. Covariates are measured in the fiscal year prior to filing for Chapter 11. *Log sales* is defined as the logarithm of total revenues in millions. *Leverage* is defined as the book value of total liabilities divided by the book value of assets. *Industry-adjusted profitability* is defined as EBITDA divided by the book value of assets, subtracted by the contemporaneous median value within each firm's three-digit NAICS industry. Log sales, leverage, and industry-adjusted profitability are winsorized at the 5% level. Standard errors are reported in brackets and clustered at the industry level.

	(1)		(2)		(3)	
	Liquidated	Acquired	Liquidated	Acquired	Liquidated	Acquired
College share	1.95	1.99	0.96	0.86	0.33	1.61
	[0.64]***	[0.68]***	[0.69]	[0.66]	[1.03]	[1.13]
Log sales			−0.11	−0.21	−0.13	−0.27
			[0.04]**	[0.05]***	[0.05]**	[0.06]***
Leverage			−1.29	−0.60	−0.91	−0.57
			[0.23]***	[0.30]**	[0.34]***	[0.44]
Industry-adjusted profitability			−0.76	0.070	−1.23	0.12
			[0.33]**	[0.41]	[0.41]***	[0.49]
Change in leverage					−0.88	0.20
					[0.47]*	[0.74]
College share X Change in leverage					1.88	−1.52
					[1.76]	[2.31]
Constant	−1.73	−1.96	1.12	−19.3	22.5	0.45
	[0.55]***	[0.63]***	[1.15]	[0.74]***	[0.72]***	[0.67]
Observations	1424		1410		1041	
Pseudo R-squared	0.026		0.059		0.070	
Log likelihood	-1329.7		-1273.9		-912.1	

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.9

Multinomial Logit Regressions Relating Post-bankruptcy Outcomes to Human Capital

The table presents estimates for the relationship between *college share* and post-bankruptcy outcomes from multinomial logit regressions of firm outcomes at the resolution of Chapter 11 for firms which filed between 1980 and 2003 and subsequently emerged as public companies. The dependent variable indicates whether the firm is continuing as an independent entity, refiled for bankruptcy, was liquidated out of court, or was acquired as of five years after emergence from Chapter 11, with continuation as the omitted category. See Section 1.3 for the definition of *college share*. Covariates are measured in the year prior to filing for Chapter 11. *Log sales* is defined as the logarithm of total revenues in millions. *Leverage* is defined as the book value of total liabilities divided by the book value of assets. *Industry-adjusted profitability* is defined as EBITDA divided the book value of assets, adjusted by its median value within a three-digit NAICS industry. Log sales, leverage, and industry-adjusted profitability are winsorized at the 5% level. Standard errors are reported in brackets and clustered at the industry level.

	(1)		(2)		(3)	
	Refiled/Liq	Acquired	Refiled/Liq	Acquired	Refiled/Liq	Acquired
College share	−4.66	−0.80	−3.56	0.14	−4.85	2.14
	[1.53]***	[1.83]	[1.41]**	[1.89]	[2.41]**	[2.29]
Log sales			0.33	0.20	0.30	0.22
			[0.09]***	[0.09]**	[0.10]***	[0.09]***
Leverage			0.25	0.91	0.068	0.48
			[0.54]	[0.49]*	[0.72]	[0.63]
Industry-adjusted profitability			−2.02	0.76	−1.82	1.14
			[0.76]***	[1.04]	[0.86]**	[1.18]
Change in leverage					−0.42	1.68
					[1.14]	[1.17]
College share X Change in leverage					3.48	−5.83
					[4.36]	[4.56]
Constant	−0.40	−1.49	−4.55	−3.95	−2.14	−3.62
	[0.59]	[0.65]**	[1.38]***	[1.21]***	[1.01]**	[1.10]***
Observations	394		391		353	
Pseudo R-squared	0.055		0.090		0.093	
Log likelihood	-336.9		-319.7		-293.9	

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.10
Multinomial Logit Regressions Relating Bankruptcy and Post-bankruptcy Outcomes to Various Measures of Human Capital

The table presents estimates from multinomial logit regressions of the relationships between *college share* and *employer tenure* and firm outcomes at the resolution of bankruptcy and after emergence from bankruptcy. See Section 1.3 for definitions of *college share* and *employer tenure*. The dependent variable in specification (1) indicates whether the firm emerged from bankruptcy as an independent entity, was liquidated within bankruptcy, or was acquired within bankruptcy, with emergence as the omitted category. The dependent variable in specification (2) indicates whether the firm was continuing as an independent entity, refiled for bankruptcy or was liquidated out of court, or was acquired as of five years after emergence from Chapter 11, with continuation as the omitted category. Covariates are measured in the year prior to filing for Chapter 11. *Log sales* is defined as the logarithm of total revenues in millions. *Leverage* is defined as the book value of total liabilities divided by the book value of assets. *Industry-adjusted profitability* is defined as EBITDA divided the book value of assets, adjusted by its median value within a three-digit NAICS industry. Log sales, leverage, and industry-adjusted profitability are winsorized at the 5% level. Marginal effects are reported at the means of the explanatory variables, with the absolute values of their Z statistics in brackets. Standard errors are clustered at the industry level.

	(1) Bankruptcy		(2) Post-bankruptcy	
	Liquidated	Acquired	Refiled/Liquidated	Acquired
College share	0.28	0.84	−4.27	−2.31
	[0.86]	[0.84]	[1.61]***	[2.20]
Employer tenure	−0.63	0.058	−0.59	−0.62
	[0.47]	[0.38]	[0.77]	[1.11]
Employees / book	−1.25	−0.0014	−0.11	−6.95
	[1.17]	[1.08]	[1.27]	[3.23]**
Log sales	−0.096	−0.2	0.31	0.24
	[0.05]**	[0.05]***	[0.09]***	[0.09]**
Leverage	−1.26	−0.58	0.3	1.31
	[0.24]***	[0.31]*	[0.52]	[0.49]***
Industry-adjusted profitability	−1.01	0.047	−2.23	0.64
	[0.34]***	[0.43]	[0.78]***	[1.01]
Constant	1.8	−19.4	−2.27	−2.49
	[1.13]	[0.74]***	[1.23]*	[1.53]
Observations	1321		362	
Pseudo R-squared	0.061		0.12	
Log likelihood	-1194.8		-286.8	

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.11
Multinomial Logit Regressions Relating Bankruptcy Outcomes to Human Capital, with Controls

The table presents estimates from multinomial logit regressions of firm outcomes at the resolution of Chapter 11 for firms which filed between 1980 and 2003. The dependent variable indicates whether the firm emerged from bankruptcy as an independent entity, was liquidated within bankruptcy, or was acquired within bankruptcy, with emergence as the omitted category. Covariates are measured in the fiscal year prior to filing for Chapter 11. *Log sales* is defined as the logarithm of total revenues in millions. *Leverage* is defined as the book value of total liabilities divided by the book value of assets. *Industry-adjusted profitability* is defined as EBITDA divided the book value of assets, adjusted by its median value within a three-digit NAICS industry. Log sales, leverage, and industry-adjusted profitability are winsorized at the 5% level. Industry medians of Q ((book value of short-term and long-term debt + market value of equity)/ book value of assets), intangibility (-PPE/book value), and R&D expenditures / book value are calculated among firms in the same three-digit NAICS industry in the Compustat database. Marginal effects are reported at the means of the explanatory variables, with the absolute values of their Z statistics in brackets. Standard errors are clustered at the industry level.

	(1)		(2)		(3)	
	Liquidated	Acquired	Liquidated	Acquired	Liquidated	Acquired
College share	0.2244	−0.0106	0.1710	0.0226	0.3183	0.0067
	[1.15]	[0.15]	[1.05]	[0.32]	[1.82]∗	[0.09]
Log sales	−0.010	−0.016	−0.013	−0.017	−0.011	−0.016
	[0.93]	[3.40]∗∗∗	[1.45]	[3.54]∗∗∗	[1.06]	[3.49]∗∗∗
Leverage	−0.275	−0.000	−0.258	−0.002	−0.281	−0.001
	[4.73]∗∗∗	[0.01]	[4.75]∗∗∗	[0.07]	[4.91]∗∗∗	[0.05]
Ind-adj profitability	−0.232	0.042	−0.194	0.041	−0.234	0.042
	[3.21]∗∗∗	[1.07]	[2.84]∗∗∗	[1.02]	[3.26]∗∗∗	[1.08]
Ind med Q	0.058	0.019				
	[1.78]∗	[1.73]∗				
Ind med intangibility			0.314	−0.000		
			[4.08]∗∗∗	[−0.01]		
Ind med R&D/book					0.029	0.071
					[0.11]	[0.83]
Constant	0.142	−0.081	0.336	−0.057	0.214	−0.057
	[1.69]∗	[1.57]	[4.36]∗∗∗	[1.23]	[2.80]∗∗∗	[1.28]
Observations	1409	1409	1409	1409	1409	1409
Log Likelihood	-1294.87		-1287.97		-1297.31	
Pseudo R2	0.04		0.05		0.04	

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 1.12
Multinomial Logit Regressions Relating Bankruptcy and Post-bankruptcy Outcomes to Human Capital, with Two-digit NAICS Fixed-effects

The table presents estimates from multinomial logit regressions of the relationships between *college share* and firm outcomes at the resolution of bankruptcy and after emergence from bankruptcy. The dependent variable in specification (1) indicates whether the firm emerged from bankruptcy as an independent entity, was liquidated within bankruptcy, or was acquired within bankruptcy, with emergence as the omitted category. The dependent variable in specification (2) indicates whether the firm was continuing as an independent entity, refiled for bankruptcy or was liquidated out of court, or was acquired as of five years after emergence from Chapter 11, with continuation as the omitted category. See Section 1.3 for definitions of *College share*. Covariates are measured in the year prior to filing for Chapter 11. *Log sales* is defined as the logarithm of total revenues in millions. *Leverage* is defined as the book value of total liabilities divided by the book value of assets. *Industry-adjusted profitability* is defined as EBITDA divided the book value of assets, adjusted by its median value within a three-digit NAICS industry. Log sales, leverage, and industry-adjusted profitability are winsorized at the 5% level. Marginal effects are reported at the means of the explanatory variables, with the absolute values of their Z statistics in brackets. Standard errors are clustered at the three-digit industry level.

	(1) Bankruptcy		(2) Post-bankruptcy	
	Liquidated	Acquired	Refiled/Liquidated	Acquired
College share	0.2124	0.0458	−0.1329	0.0001
	[1.39]	[0.89]	[2.83]***	[0.00]
Log sales	−0.019	−0.012	0.01	0.005
	[2.14]**	[3.40]***	[3.31]***	[2.12]**
Leverage	−0.257	−0.014	0.015	0.03
	[4.73]***	[0.60]	[0.83]	[2.05]**
Ind-adj profitability	−0.167	0.02	−0.057	0.022
	[2.34]**	[0.66]	[2.33]**	[0.92]
Constant	0.802	−1.658	−0.052	−0.682
	[3.04]***	[11.85]***	[1.15]	[4.81]***
Observations	1409	1409	391	391
Log Likelihood	-1248.56		-301.34	
Pseudo R2	0.08		0.14	
2-digit NAICS FE	yes	yes	yes	yes

* significant at 10%; ** significant at 5%; *** significant at 1%

Chapter 2

Superstar Extinction Coauthored with Pierre Azoulay and Joshua Graff Zivin

2.1 Introduction

Although the production of ideas occupies a central role in modern theories of economic growth (Romer 1990), the creative process remains a black box for economists (Weitzman 1998 and Jones 2009 are notable exceptions). How do innovators actually generate new ideas? Increasingly, discoveries result from the voluntary sharing of knowledge through collaboration, rather than individual efforts (Wuchty et al. 2007). The growth of scientific collaboration has important implications for the optimal allocation of public R&D funds, the apportionment of credit amongst scientists, the formation of scientific reputations, and ultimately the design of research incentives that foster innovation and continued economic growth. Yet, we know surprisingly little about the role of collaboration among peers as a mechanism to spur the creation of new technological or scientific knowledge.

This paucity of evidence is largely due to the empirical challenges inherent to this line of inquiry. Individual-level data on the contributors to a particular innovation are generally unavailable. Furthermore, the formation of collaborative teams is the outcome of a purposeful matching process (Mairesse and Turner 2005; Fafchamps et al. 2008), making it difficult to uncover causal effects. The design of our study tackles both of these challenges. To relax the data constraint, we focus on the academic life sciences, where a rich tradition of coauthorship provides an extensive paper trail of collaboration histories and research output. To overcome the endogeneity of the collaboration decision, we make use of the quasi-experimental variation in the structure of coauthorship networks induced by the premature and sudden death of active "superstar" scientists.[1]

Specifically, we analyze changes in the research output of collaborators for 112 eminent life scientists who die suddenly and unexpectedly. We assess eminence based on the combination of seven criteria, and our procedure is flexible enough to capture established scientists with extraordinary career achievement, as well as promising young and mid-career scientists. Using the Association of American Medical Colleges (AAMC) Faculty Roster as a data source — a comprehensive, longitudinal, matched employee-employer database pertaining to 230,000 faculty members in all U.S. medical schools between 1975 and 2006 — we construct a panel dataset of 5,267 collaborator-star pairs, and we examine how coauthors' scientific output (as measured by publications, citations, and National Institutes of Health (NIH) grants) changes when the superstar passes away.[2]

The study's focus on the scientific elite can be justified both on substantive and pragmatic grounds. The distribution of publications, funding, and citations at the individual level is extremely skewed (Lotka 1926; de Solla Price 1963) and only a tiny minority of scientists contribute through their published research to the advancement of science (Cole and Cole 1972). Stars also leave behind a corpus of work and colleagues with a stake in the preservation of their legacy, making it possible to trace back their careers, from humble beginnings to wide recognition and acclaim.

Our results reveal a lasting 5 to 8% decrease in the quality-adjusted publication output of coauthors in response to the sudden and unexpected loss of a superstar. Though close and recent collaborators see their scientific output fall even more, these differential effects are small in magnitude and statistically insignificant. Therefore, the process of replacing missing skills within ongoing collaborative teams cannot, on its own, explain our core result.

The importance of learning through on-the-job social interactions can be traced back to the talmudic era (as evidenced by the epigraph to this paper), as well as canonical writings by Alfred Marshall (1890) and Robert Lucas (1988).[3] Should the effects of exposure to superstar talent be interpreted as laying bare the presence of knowledge spillovers? Since we identify 47 coauthors per superstar on average, we exploit rich variation in the characteristics of collaborative relationships to assess the relative importance of several mechanisms which could plausibly account for our main finding.

[1]Other economists have used the death of prominent individuals as a source of exogenous variation in leadership, whether in the context of business firms (Bennedsen et al. 2008), or even entire countries (Jones and Olken 2005). To our knowledge, however, we are the first to use this strategy to estimate the impact of scientific collaboration. Oettl (2008) builds on our approach by incorporating helpfulness as implied by acknowledgements to generate a list of eminent immunologists. Aizenman and Kletzer (2008) study the citation "afterlife" of 16 economists who die prematurely, shedding light on the survival of scientific reputation.

[2]To be clear, our focus is on faculty peers rather than trainees, and thus our results should be viewed as capturing inter-laboratory spillovers rather than mentorship effects. For evidence on the latter, see Azoulay et al. (2009).

[3]A burgeoning empirical literature examines the influence of peer effects on shirking behavior in the workplace (Costa and Khan 2003; Bandiera et al. 2005; Mas and Moretti 2009). Since "exposure" does not involve the transmission of knowledge, these spillovers are conceptually distinct from those that concern us here.

A jaundiced view of the academic reward system provides the backdrop for a broad class of stories. Their common thread is that collaborating with superstars deepens social connections that might make researchers more productive in ways that have little to do with scientific knowledge, for example by connecting coauthors to funding resources, editorial goodwill, or potential coauthors. Yet, we find no differential impact on coauthors of stars well-connected to the NIH funding apparatus, on coauthors of stars more central in the collaboration network, or on former trainees. These findings do not jibe with explanations stressing the gatekeeping role of eminent scientists.

Rather, the effects of superstar extinction appear to be driven by the loss of an irreplaceable source of ideas. We find that coauthors proximate to the star in intellectual space experience a sharper decline in output, relative to coauthors who work on less related topics. Furthermore, the collaborators of stars whose work was heavily cited at the time of their death also undergo steeper decreases, relative to collaborators of superstars of less renown. Together, these results paint a picture of an invisible college of coauthors bound together by interests in a fairly specific scientific area, which suffers a permanent and reverberating intellectual loss when it loses its star.

The rest of the paper proceeds as follows. In the next section, we describe the construction of the sample of matched superstars and collaborators, as well as our empirical strategy. Section 3 provides descriptive statistics at the coauthor and dyad level. We report the results in section 4. Section 5 concludes.

2.2 Setting, Data, and Matched Sample Construction

The setting for our empirical work is the academic life sciences. This sector is an important one to study for several reasons. First, there are large public subsidies for biomedical research in the United States. With an annual budget of $29.5 billion in 2008, support for the NIH dwarfs that of other national funding agencies in developed countries (Cech 2005). Deepening our understanding of knowledge production in this sector will allow us to better assess the return to these public investments.

Second, technological change has been enormously important in the growth of the health care economy, which accounts for roughly 15% of US GDP. Much biomedical innovation is science-based (Henderson et al. 1999), and interactions between academic researchers and their counterparts in industry appear to be an important determinant of research productivity in the pharmaceutical industry (Cockburn and Henderson 1998; Zucker et al. 1998).

Third, academic scientists are generally paid through soft money contracts. Salaries depend on the amount of grant revenue raised by faculty, thus providing researchers with high-powered incentives to remain productive even after they secure a tenured position.

Lastly, introspective accounts by practicing scientists indicate that collaboration plays a large role in both the creation and diffusion of new ideas (Reese 2004). Knowledge and techniques often remain partially tacit until long after their initial discovery, and are transmitted within the confines of tightly-knit research teams (Zucker and Darby 2008).

2.2.1 Superstar Sample

Our basic approach is to rely on the death of "superstar" scientists to estimate the magnitude of knowledge spillovers onto colleagues. From a practical standpoint, it is more feasible to trace back the careers of eminent scientists than to perform a similar exercise for less eminent ones. We began by delineating a set of 10,349 "elite" life scientists (roughly 5% of the entire relevant labor market) who are so classified if they satisfy at least one of the following criteria for cumulative scientific achievement: (1) highly funded scientists; (2) highly cited scientists; (3) top patenters; and (4) members of the National Academy of Sciences.

These four criteria will tend to select seasoned scientists, since they correspond to extraordinary achievement over an entire scientific career. We combine these measures with three others that capture individuals who show great promise at the early and middle stages of their scientific careers, whether or not these episodes of productivity endure for long periods of time: (5) NIH MERIT awardees; (6) Howard Hughes Medical Investigators; and (7) early career prize winners. Appendix I provides additional details regarding these seven metrics of "superstardom."

We trace back these scientists' careers from the time they obtain their first position as independent investigators (typically after a postdoctoral fellowship) until 2006. We do so through a combination of curriculum vitæs, NIH biosketches, *Who's Who* profiles, accolades/obituaries in medical journals, National Academy of Sciences biographical memoirs, and Google searches. For each one of these individuals, we record employment history, degree held, date of degree, gender, and up to three departmental affiliations. We also cross-reference the list with alternative measures of scientific eminence. For example, the elite subsample contains every U.S.-based Nobel Prize winner in Medicine and Physiology since 1975, and a plurality of the Nobel Prize winners in Chemistry over the same time period.

Though we apply the convenient moniker of "superstar" to the entire group, it should be clear that there is substantial heterogeneity in intellectual stature within the elite sample. This variation provides a unique opportunity to examine whether the effects we estimate correspond to vertical effects (spillovers from the most talented agents onto those who are less distinguished) rather than peer effects (spillovers between agents of roughly comparable stature).

The scientists who are the focus of this paper constitute a subset of this larger pool of 10,349. We impose several additional criteria to derive the final list. First, the scientist's death must intervene between 1979 and 2003. This will enable us to observe at least 4 years' (resp. 3 years') worth of scientific output for every colleague before (resp. after) the death of their superstar collaborator. Second, they must be 67 years of age or less at the time of their passing (we will explore the sensitivity of our results to this age cutoff later). Third, we require evidence, in the form of published articles and/or NIH grants, that these scientists have not entered a pre-retirement phase of their career prior to the time of death. This constraint is somewhat subjective, but we validate in the on-line appendix our contention that the final set is limited to scientists who are "research-active" at the time of their death. These sequential screens delineate a set of 248 scientists. Finally, we limit our attention to the subset of stars who died suddenly and unexpectedly. This is less difficult that it might seem, since the vast majority of obituaries mention the cause of death explicitly.[4] After eliminating 136 scientists whose death could have been anticipated by their colleagues, we are left with 112 extinct superstars (their names, cause of death, and institutional affiliations are listed in Table W1 in the on-line appendix).

Table I provides descriptive statistics for the superstar sample. The average star received his degree in 1963, died at 57 years old and worked with 47 coauthors during his lifetime. On the output side, the stars each received an average of roughly 11 million dollars in NIH grants (excluding center grants), and published 139 papers that garnered 8,190 citations as of early 2008.

2.2.2 The Universe of Potential Colleagues

Information about the superstars' colleagues stems from the Faculty Roster of the Association of American Medical Colleges, to which we secured licensed access for the years 1975 through 2006. The roster is an annual census of all U.S. medical school faculty in which each faculty is linked across yearly cross-sections by a unique identifier.[5] When all cross-sections are pooled, we obtain a matched employee/employer panel dataset. For each of the 230,000 faculty members that appear in the roster, we know the full name, the type of degrees received and the years they were awarded, gender, up to two departments, and medical school affiliation. An important implication of our reliance on the AAMC Faculty Roster is that the interactions we can observe in the data take place between faculty members, rather than between faculty members and trainees (graduate students or post-doctoral fellows).[6]

Because the roster only lists medical school faculty, however, it is not a complete census of the academic life sciences. For instance, it does not list information for faculty at institutions

[4] We exclude from the sample one scientist who took his own life, and a further two for whom suicide could not be ruled out. In 10 other instances, the cause of death could not be ascertained from the obituaries and we contacted former collaborators individually to clarify the circumstances of the superstar's passing.

[5] AAMC does not collect data from each medical school with a fixed due date. Instead, it collects data on a rolling basis, with each medical school submitting on a time frame that best meets its reporting needs. Nearly all medical schools report once a year, while many medical schools update once a semester.

[6] To the extent that former trainees go on to secure faculty positions, they will be captured by our procedure even if the date of coauthorship predates the start of their independent career.

such as MIT, University of California at Berkeley, Rockefeller University, the Salk Institute, or the Bethesda campus of the NIH; and it also ignores faculty members in Arts and Sciences departments — such as biology and chemistry — if they do not hold joint appointments at a local medical school.[7]

Our interest lies in assessing the benefits of exposure to superstar talent that accrue through collaboration. Therefore, we focus on the one-degree, egocentric coauthorship network for the sample of 112 extinct superstars. To identify coauthors, we have developed a software program, the Stars/Colleague Generator, or S/CGEN.[8] The source of the publication data is PubMED, an online resource from the National Library of Medicine that provides fast, free, and reliable access to the biomedical research literature. In a first step, S/CGEN downloads from the internet the entire set of English-language articles for a superstar, provided they are not letters to the editor, comments, or other "atypical" articles. From this set of publications, S/CGEN strips out the list of coauthors, eliminates duplicate names, matches each coauthor with the Faculty Roster, and stores the identifier of every coauthor for whom a match is found. In a final step, the software queries PubMED for each validated coauthor, and generates publication counts as well as coauthorship variables for each superstar/colleague dyad, in each year. In the on-line appendix, we provide details on the matching procedure, how we guard against the inclusion of spurious coauthors, and our approach to addressing measurement error when tallying the publication output of coauthors with common names.

2.2.3 Identification Strategy

A natural starting point to identify the effect of superstar death is to examine changes in collaborator research output after the superstar passes away, relative to when s/he was still alive, using a simple collaborator fixed effects specification. Since the extinction effect is mechanically correlated with the passage of time, as well as with coauthor's age, our specifications must include life cycle and period effects, as is the norm in studies of scientific productivity (Levin and Stephan 1991). In this framework, the control group that pins down the counterfactual age and calendar time effects for the coauthors that currently experience the death of a superstar consists of coau-

[7]This limitation is less important than might appear at first glance. First, we have no reason to think that colleagues located in these institutions differ in substantive ways from those based in medical schools. Second, all our analyses focus on *changes* in research productivity over time for a given scientist. Therefore, the limited coverage is an issue solely for the small number of faculty who transition in and out of medical schools from (or to) other types of research employment. For these faculty, we were successful in filling career gaps by combining the AAMC Faculty Roster with the NIH data.

[8]The software can be used by other researchers under an open-source (GNU) license. It can be downloaded, and detailed specifications accessed, from the web site `http://stellman-greene.com/SCGen/`. Note that the S/CGEN takes the AAMC Faculty Roster as an input; we are not authorized to share this data with third-parties. However, it can be licensed from AAMC, provided a local IRB gives its approval and a confidentiality agreement protects the anonymity of individual faculty members.

thors whose associated superstar died in earlier periods, or will die in future periods. Despite its long pedigree in applied economics (e.g., Grogger 1995; Reber 2006), this approach may be problematic in our setting.

First, coauthors observed in periods after the death of their associated superstar are not appropriate controls if the event negatively affected the trend in their output; if this is the case, fixed effects will underestimate the true effect of superstar extinction. Second, collaborations might be subject to idiosyncratic life cycle patterns, with their productive potential first increasing over time, eventually peaking, and thereafter slowly declining; if this is the case, fixed effects will overestimate the true effect of superstar extinction, at least if we rely on collaborators treated in earlier or later periods as as an "implicit" control group.

To mitigate these threats to identification, our preferred empirical strategy relies on the selection of a matched control for each scientist who experiences the death of a superstar collaborator. These control scientists are culled from the universe of coauthors for the 10,000 superstars who do not die (see Section 2.1). Combining the treated and control samples enables us to estimate the effect of superstar extinction in a difference in differences framework. Using a "coarsened exact matching" procedure detailed in Appendix II, the control coauthors are chosen so that (1) treated scientists exhibit no differential output trends relative to controls up to the time of superstar death; (2) the distributions of career age at the time of death are similar for treated and controls; (3) the time paths of output for treated and control coauthors are similar up to the time of death; and (4) the dynamics and main features of collaboration (number of coauthorships at the time of death, time elapsed since first and last coauthorship; status of the superstar collaborator as summarized by cumulative citations in the year of death) are balanced between treated and controls.

However, adding this control group to the basic regression does not, by itself, yield a specification where the control group consists *exclusively* of matched controls. Figure A1 displays the trends in average and median number of quality-adjusted publications, for treated and control collaborators respectively, without any adjustment for age or calendar time effects. This raw comparison is not without its problems, since it involves centering the raw data around the time of death, thus ignoring the lack of congruence between experimental and calendar time. Yet, it is completely non-parametric, and provides early evidence that the loss of a superstar coauthor leads to a decrease in collaborators' publication output. Furthermore, the magnitude of the estimates presented below are very similar whether or not control scientists are added to the estimation sample.

Another potential concern with the addition of this "explicit" control group is that control coauthors could be affected by the treatment of interest. No scientist is an island. The set of coauthors for our 10,349 elite scientists comprises 65% of the labor market, and the remaining 35% corresponds in large part to clinicians who hold faculty appointments but do not publish regularly.

Furthermore, the death of a prominent scientist could affect the productivity of non-coauthors if meaningful interactions take place in "idea space," as we propose. Thus, in robustness checks, we check whether eliminating from the estimation sample treated and control collaborators separated by small path lengths in the coauthorship network matters for the substance, or even the magnitudes, of our main results.

2.3 Descriptive Statistics

When applied to our sample of 112 extinct superstars, S/CGEN identifies 5,267 distinct coauthors with unique PubMED names.[9] Our matching procedure can identify a control scientist for 5,064 (96%) of the treated collaborators. The descriptive statistics in Table II pertain to the set of $2 \times 5,064 = 10,128$ matched treated and control scientists. The covariates of interest are measured in the year of the (possibly counterfactual) year of death for the superstar. We distinguish between variables that are inherently dyadic (e.g., co-location at time of death) from variables that characterize the coauthor at a particular point of time (e.g., NIH R01 funding at the time of death).

Dyadic variables. Of immediate interest is the distribution of coauthorship intensity at the dyad level. While the average number of coauthorships is slightly less than three, the distribution is extremely skewed (Figure I). We define "casual" dyads as those that have two or fewer coauthorships with the star, "regular" dyads as those with three to nine coauthorships, and "close" dyads as those with ten or more coauthorships. Using these cutoffs, "regular" dyads correspond to those between the 75^{th} and the 95^{th} percentile of coauthorship intensity, while "close" dyads correspond to those above the 95^{th} percentile.

We focus next on collaboration age and recency. On average, collaborations begin 11 years before the star's death, and time since last coauthorship is slightly more than 9 years. In other words, most of the collaborations in the sample do not involve active research projects at the time of death. Recent collaborations (those that involve at least one coauthorship in the three years preceding the passing of the superstar) map into the top quartile of collaboration recency at the dyad level.

The research collaborations studied here occur between faculty members, who often run their own labs (a conjecture reinforced by the large proportion of coauthors with independent NIH funding). Yet, it is interesting to distinguish collaborators who trained under a superstar (either in graduate school or during a postdoctoral fellowship) from those collaborations initiated at a time

[9]Whenever a scientist collaborates with more than one extinct superstar (this is relevant for 10% of the sample), we take into account only the first death event. We have verified that limiting the estimation sample to collaborators with one and only one tie to a superstar who dies does not change the substance, or even the magnitudes, of our core result.

in which both nodes in the dyad already had a faculty appointment. While there is no roster of mentor/mentee pairs, coauthorship norms in the life sciences provide an opportunity to identify former trainees. Specifically, we flag first-authored articles published within a few years of receipt of the coauthor's degree in which the superstar appears in last position on the authorship roster.[10] Using this method, we find that roughly 8% of treated collaborators were former trainees of the associated superstar.

We now examine the spatial distribution of collaborations. Slightly more than 12% of collaborations correspond to scientists who were co-located at the time of superstar extinction; though this is not the focus of the paper, the proportion of local collaborations has declined over time, as many previous authors have documented (e.g., Rosenblat and Möbius 2004). We also provide a measure of collaborators' proximity in "ideas space." Every publication indexed by PubMED is tagged by a large number of descriptors, selected from a dictionary of approximately 25,000 MeSH (*Me*dical *S*ubject *H*eadings) terms. Our measure of intellectual proximity between members of a dyad is simply the number of unique MeSH terms which overlap in their non-coauthored publications, normalized by the total number of MeSH terms used by the superstar's coauthor. The time window for the calculation is the five years that precede the passing of the superstar. The distribution of this variable is displayed in Figure II.[11]

Finally, we create a measure of social proximity that relies not on the quantity of coauthored output, but on the degree of social interaction it implies. We focus on the pairs involving coauthors who, whenever they collaborate, find themselves in the middle of the authorship list. Given the norms that govern the allocation of credit in the life sciences, these coauthors are likely to share the least amount of social contact. 7.5% of the dyads in the sample correspond to this situation of "accidental coauthorship" — the most tenuous form of collaboration.

Coauthor variables. We briefly mention demographic characteristics that do not play a role in the econometric results but are nonetheless informative. The sample is 20% female (only 10% of the superstars are women); approximately half of all coauthors are MDs, 40% are PhDs, and the remainder are MD/PhDs; and a third are affiliated with basic science departments (as opposed to clinical or public health departments). The coauthors are about 8 years younger than the superstars on average (1971 vs. 1963 for the year of highest degree).

Coauthors lag behind superstars in terms of publication output at the time of death, but the

[10]The purported training period runs from 3 years before graduation to 4 years after graduation for PhDs and MD/PhDs; and from the year of graduation to 6 years after graduation for MDs. Recall that we do not observe the population of former trainees, but only those trainees that subsequently went on to get full-time faculty positions in the United States. One concern is selection bias for the set of former trainees associated with superstars who died when they had just completed training. To guard against this potential source of bias, we eliminated all former trainees from the sample with career age less than 5 at the time of death.

[11]Further details on its construction are provided in the on-line appendix, Section II.

difference is not dramatic (88 vs. 140 articles, on average). Assortative matching is present in the market for collaborators, as reflected by the fact that 2,852 (28.16%) of our 10,128 coauthors belong to the elite sample of 10,349 scientists. 55% of collaborators had served as PI on at least one NIH R01 grant when the superstar passes away, while about 8% of the treated collaborators (and 9% of the controls) belong to a more exclusive elite: Howard Hughes Medical Investigators, members of the NAS, or MERIT awardees.

The estimation sample pools observations between 1975 and 2006 for the dyads described above. The result is an unbalanced panel dataset with 153,508 collaborator×year observations (treated collaborators only) or 294,943 collaborator×year observations (treated and control collaborators).

2.4 Results

The exposition of the econometric results proceeds in three stages. After a brief review of methodological issues, we provide results that pertain to the main effect of superstar exposure on publication rates. Second, we examine whether this effect merely reflects the adverse impact of losing important skills within ongoing collaborative teams. Third, we attempt to explicate the mechanism, or set of mechanisms, responsible for the results. We do so by exploring heterogeneity in the treatment through the interaction of the post-death indicator variable below with various attributes of the superstar, colleague, and dyad.

2.4.1 Econometric Considerations

Our estimating equation relates colleague j's output in year t to characteristics of j, superstar i, and dyad ij:

$$E\left[y_{jt}|X_{ijt}\right] = exp\left[\beta_0 + \beta_1 AFTER\_DEATH_{it} + f(AGE_{jt}) + \delta_t + \gamma_{ij}\right] \quad (2.1)$$

where y is a measure of research output, $AFTER\ DEATH$ denotes an indicator variable that switches to one the year after the superstar dies, $f(AGE_{jt})$ corresponds to a flexible function of the colleague's career age, the δ_t's stand for a full set of calendar year indicator variables, and the γ_{ij}'s correspond to dyad fixed effects, consistent with our approach to analyze *changes* in j's output following the passing of superstar i.

The dyad fixed effects control for many individual characteristics that could influence research output, such as gender or degree. Academic incentives depend on the career stage; given the shallow slope of post-tenure salary increases, Levin and Stephan (1991) suggest that levels of investment in research should vary over the career life cycle. To flexibly account for life cycle effects, we include seventeen indicator variables corresponding to different career age brackets, where career age measures the number of years since a scientist earned his/her highest degree (MD or PhD).[12] In specifications that include an interaction between the treatment effect and some covariates, the models also include a set of interactions between the life cycle effects and these covariates.

Estimation. The dependent variables of interest, including weighted or unweighted publication counts and NIH grants awarded, are skewed and non-negative. For example, 24.80% of the collaborator/year observations in the data correspond to years of no publication output; the figure climbs to 87.40% if one focuses on the count of successful grant applications. Following a long-standing tradition in the study of scientific and technical change, we present conditional quasi-maximum likelihood estimates based on the fixed-effect Poisson model developed by Hausman et al. (1984). Because the Poisson model is in the linear exponential family, the coefficient estimates remain consistent as long as the mean of the dependent variable is correctly specified (Gouriéroux et al. 1984).[13]

Inference. QML (i.e., "robust") standard errors are consistent even if the underlying data generating process is not Poisson. In fact the Hausman et al. estimator can be used for any non-negative dependent variables, whether integer or continuous (Santos Silva and Tenreyro 2006), as long as the variance/covariance matrix is computed using the outer product of the gradient vector (and therefore does not rely on the Poisson variance assumption). Further, QML standard errors are robust to arbitrary patterns of serial correlation (Wooldridge 1997), and hence immune to the issues highlighted by Bertrand et al. (2004) concerning inference in DD estimation. We cluster the standard errors around superstar scientists in the results presented below.

Dependent Variables. Our primary outcome variable is a coauthor's number of publications. Since SC/GEN matches the entire authorship roster for each article, we can separate those publications coauthored with the superstar from those produced independently of him/her. We perform a quality adjustment by weighting each publication by its Journal Impact Factor (JIF) — a measure of the frequency with which the "average article" in a journal has been cited in a particular year. One obvious shortcoming of this adjustment is that it does not account for differences in impact within a given journal. In the on-line appendix (section V), we present additional results based on

[12]The omitted category corresponds to faculty members in the very early years of their careers (before age -3). It is not possible to separately identify calendar year effects from age effects in the "within" dimension of a panel in a completely flexible fashion, because one cannot observe two individuals at the same point in time that have the same (career) age but earned their degrees in different years (Hall et al. 2007).

[13]In the on-line appendix (section IV), we show that OLS yields very similar results to QML Poisson estimation for our main findings.

article-level citation outcomes.

2.4.2 Main effect of superstar extinction

Table III presents our core results. Column 1a examines the determinants of the 5,267 treated coauthors' JIF-weighted publication output. We find a sizable and significant 8.8% decrease in the yearly number of quality-adjusted publications coauthors produce after the star dies. Column 2b adds the set of control coauthors to the estimation sample. This reduces only slightly our estimate of the treatment effect, to a statistically significant 8.2% decline.

Columns 1b and 2b provide the results for an identical set of specifications, except that we modify the dependent variable to exclude publications coauthored with the superstar when computing the JIF-weighted publication counts. The contrast between the results in Panels A and B elucidates scientists' ability to substitute towards new collaborative relationships upon the death of their superstar coauthor. The effects are now smaller, but they remain statistically significant.

We also explore the dynamics of the effects uncovered in Table III. We do so by estimating a specification in which the treatment effect is interacted with a set of indicator variables corresponding to a particular year relative to the superstar's death, and then graphing the effects and the 95% confidence interval around them (Figures IIIA and IIIB, corresponding to Table III, columns 1b and 2b). Following the superstar's death, the treatment effect increases monotonically in absolute value, becoming statistically significant three to four years after death. Two aspects of this result are worthy of note. First, we find no evidence of recovery — the effect of superstar extinction appears permanent. Though we will explore mechanisms in more detail below, this seems inconsistent with a bereavement-induced loss in productivity. Second, the delayed onset of the effect makes sense because it plausibly takes some time to exhaust the productive potential of the star's last scientific insights. In addition, the typical NIH grant cycle is three to five years, and the impact of a superstar's absence may not really be felt until it becomes time to apply for a new grant.

In all specifications, the results with and without controls are quite similar. In the remainder of the paper, the estimations sample always include the "explicit" control group, though the results without it are qualitatively similar.

2.4.3 Imperfect Skill Substitution

Collaborative research teams emerge to pool the expertise of scientists, who, in their individual capacity, face the "burden of knowledge" problem identified by Jones (2009). Upon the death of a key collaborator, other team members might struggle to suitably replace the pieces of knowledge that were embodied in the star. Viewed in this light, the effects uncovered in Table III could be considered unsurprising — a mechanical reflection of the skill substitution process. The fact that publications with coauthors other than the superstar are adversely affected, and the permanence of the treatment effect already suggest other forces are at play. The imperfect skill substitution (ISS) story carries additional testable implications. First, one would expect coauthors with closer relationships with the star to suffer steeper decreases in output; the same would be expected for recent or new collaborations, which are more likely to involve ongoing research efforts at the time of death. Table IV examines these implications empirically.

We find that regular, and to a lesser extent, close collaborators are indeed more negatively affected than casual collaborators, but these differential losses are relatively small in magnitude and statistically insignificant (column 1a). The same holds true for recent collaborations (at least one joint publication in the three years preceding the star's death, column 2a) and for "young" collaborations (those for which the first coauthored publication appeared in the five years preceding the star's death, unreported results available from the authors). Columns 1b and 2b provide results for an identical set of specifications, but excluding publications coauthored with the superstar. The contrast between the results in columns 1a and 1b (resp. 2a and 2b) elucidates scientists' ability to substitute towards new collaborative relationships upon the death of their superstar coauthor. The estimates imply that close and, to a lesser extent, recent coauthors do manage to find replacement collaborators (or to intensify already existing collaborations). Close collaborators experience an imprecisely estimated 6.18% average *increase* in their quality-adjusted publications written independently of the star, but this is only a partial offset for the overall loss documented in column 1a. We find that casual collaborators and collaborators without a recent coauthorship see their independent output decline respectively by 5.54% (column 1b) and 8.25% (column 2b). Very similar results are obtained when combining all these covariates into one specification (columns 3a and 3b).

While the differential impacts on the closest and most recent collaborators are not statistically significant, they do appear to move in the direction that supports the skill substitution hypothesis. However, the inability of scientists to fully compensate for the loss of expected future collaborations through alternative relationships, as well as the permanence of the extinction effect, demonstrate that something more than the star's skills disappears upon their death. Taken as a whole, these results suggest that the treatment effect from Table III cannot be fully explained by imperfect skill substitution within ongoing teams.

2.4.4 Disentangling Mechanisms

We exploit the fine-grained level of detail in the data to sort between mechanisms which might underlie the extinction effect. Are collaborative ties with superstars conduits for tangible resources, or for knowledge and ideas? These two broad classes of explanations are not mutually exclusive, but ascertaining their relative importance matters because their welfare implications differ sharply. If superstars merely act as gatekeepers, then their deaths will lead to a reallocation of resources away from former collaborators, but may have little impact on social welfare. Conversely, if spillovers of knowledge were enabled by collaboration, their passing might result in significant welfare losses.

Superstars as Gatekeepers. Superstars may matter for their coauthors because they connect them to important resources either within their institution or in the scientific world at large. These resources might include funding, administrative clout, editorial goodwill, or other potential collaborators. We attempt to evaluate the validity of three particular implications of this story in Table V.

First, we examine whether the superstar's ties to the NIH funding apparatus moderate the magnitude of the extinction effect. Whereas social scientists sometimes emphasize the role that journal editors can have in shaping individual careers, life scientists are often more concerned that the allocation of grant dollars deviates from the meritocratic ideal. Therefore, we investigate whether the treatment effect is of larger magnitude when the star either sat on NIH review panels in the last 5 years, or has coauthorship ties with other scientists who sat on study sections in the recent past. In column 1, we find that this is not the case. The differential impacts are relatively small, positive in magnitude, and not statistically significant.

Second, we address the hypothesis that superstars matter because they broker relationships between scientists that would otherwise remain unaware of each other's expertise. We do so by computing the betweenness centrality for the extinct superstars in the coauthorship network formed by the 10,349 elite scientists.[14] We then rank the superstars according to quartile of betweenness, and look for evidence that collaborators experience a more pronounced decline in output if their superstar coauthor was more central (column 2). We find that collaborators with stars in the top quartile suffer additional losses, relative to collaborators of less central superstars, but this differential effect is statistically insignificant.

Finally, in column 3, we look for a differential effect of superstar death for coauthors that were also former trainees. It is possible that mentors continue to channel resources to their former associates even after they leave their labs, in which case one would expect these former trainees to exhibit steeper and more precipitous declines following the passing of their adviser. In fact, the differential effect is large and positive, though not statistically significant.

The evidence presented in Table V appears broadly inconsistent with the three particular gatekeeping stories whose implications we could test empirically. Our assessment of the gatekeeping mechanism must remain guarded for two reasons. First, the effect of variables used to proxy for the strength of social ties are subject to alternative interpretations. For instance, a former trainee

[14]Betweenness is a measure of the centrality of a node in a network, and is calculated as the fraction of shortest paths between dyads that pass through the node of interest. In social network analysis, it is often interpreted as a measure of the influence a node has over the spread of information through the network.

effect could also be interpreted as providing evidence of knowledge spillovers, since mentorship can continue into the early faculty career and be extremely important for a young scholar's intellectual development. Furthermore, it is possible to think of alternative versions of the gatekeeping mechanism; as an example, superstars might be able to curry favors with journal editors on behalf of their protégés, or they might be editors themselves. We prefer to frame the findings contrapositively: it is hard to look at the evidence presented so far and conclude that access to resources is a potent way in which superstars influence their collaborators' scientific output.

Knowledge Spillovers. We now examine the possibility that stars generate knowledge spillovers onto their coauthors. In Table VI, we build a circumstantial case for the spillover view by documenting evidence of additional output losses for collaborators who were more proximate with the superstar at the time of death, using two different meanings of proximity: physical and intellectual.

In column 1, we investigate the impact of physical proximity by interacting the treatment effect with an indicator variable for those collaborators who were co-located with the superstar at the time of death. We find essentially no difference between the fates of these coauthors and those of coauthors located further away — the interaction term is positive, small in magnitude, and imprecisely estimated. At first blush, this finding appears consistent with some recent work suggesting a fading role for geographic distance, both as a factor influencing the formation of teams (Rosenblat and Möbius 2004; Agrawal and Goldfarb 2008), and as a factor circumscribing the influence of peers (Kim et al. 2006; Griffith et al. 2007; Waldinger 2008). However, our estimate of the co-location interaction term conflates the effect of the loss of knowledge spillovers, the effect of the loss of help and protection provided by the star in the competition for internal resources (such as laboratory space), and the effect of any measure taken by the institution to compensate for the death of the superstar. As a result, it is unclear whether our results contradict the more conventional view that spillovers of knowledge are geographically localized (Zucker et al. 1998; Ham and Weinberg 2008).[15]

In column 2, we investigate whether the death of a superstar coauthor has a disparate impact on the group of scientists who work on similar research problems. We proxy intellectual distance between the superstar and his/her coauthors with our measure of normalized keyword overlap. Coauthors in the top quartile of this measure at the time of death suffer output decreases that are particularly large in magnitude (-12.2%).[16] This evidence is consistent with the existence of an "invisible college" — an elite of productive scientists highly visible in a research area, combined with a "scatter" of less eminent ones, whose attachment to the field may be more tenuous (de

[15]We thank an anonymous reviewer for making this point.

[16]Specifications that include four different interactions corresponding to the four quartiles show that the treatment effect is monotonically increasing in intellectual distance, but we do not have enough statistical power to reject the hypothesis that the five coefficients are equal to one another.

Solla Price and Beaver 1966; Crane 1972). Superstar scientists make their field of inquiry visible to others of lesser standing who might enter it; they replenish their field with fresh ideas, and their passing causes the processes of knowledge accumulation and diffusion to slow down, or even decline.

In this view, important interactions for the production of new scientific knowledge are not rigidly constrained by geographic or social space, but also take place in an ethereal "idea space." But is the act of formal coauthorship necessary for a scientist to be brought into a superstar's intellectual orbit? Since our sample is composed exclusively of coauthors, we cannot definitively answer this question. Yet, one can use the norms of authorship in the life sciences to try to isolate collaborators whose coauthorship tie to the star is particularly tenuous: "accidental" collaborators — those who always find themselves in the middle of the authorship list. As seen in column 3, these accidental collaborators do not appear to experience net losses after the superstar's death. This suggests that full membership in the invisible college may be difficult to secure in the absence of a preexisting social tie.

Column 4 provides evidence that the effects of physical and intellectual proximity are independent, since combining them in the same specification does not alter their magnitudes or statistical significance. Finally, column 5 demonstrates that these effects are robust to the inclusion of controls for coauthorship intensity and recency.

Table VII provides additional evidence in favor of the spillover view by examining the relationship between the magnitude of the treatment effect and the accomplishments of the star. We rank superstars according to two metrics of achievement: cumulative citations and cumulative NIH funding, and we focus on superstars in the top quartile of either distributions (where these quartiles are calculated using the population of 10,349 superstars in a given year). Column (1) shows that collaborators of heavily cited superstars suffer more following the superstar's death, while column (2) shows that this is not true for collaborators of especially well-funded superstars. Column (3) puts the two effects in a single specification. Once again, it appears that it is the star's citation impact that matters in shaping collaborators' post-extinction outcomes, rather than his/her control over a funding empire.[17] We interpret these findings as buttressing our argument that it is the quality of ideas emanating from the stars, rather than simply the availability of the research funding they control, that goes missing after their deaths. Furthermore, these results suggest that using the same empirical strategy, but applying it to a sample of "humdrum" coauthors who die, would not uncover effects similar in magnitude to those we observed in Table III. As such, they validate *ex post* our pragmatic focus on the effect of superstars.

The overall collection of results presented above help build a circumstantial case in favor of

[17]Table VII eliminates from the estimation sample the collaborators of 11 superstars who are NIH intramural scientists, and as such not eligible for extramural NIH funding.

interpreting the effects of superstar extinction as evidence of missing spillovers. However, they do not enable us to reject some potentially relevant versions of the gatekeeping story — such as influence over the editorial process in important journals, nor do they allow us to learn about the effect on non-collaborators.

2.4.5 Robustness and Sensitivity Checks

The on-line appendix provides additional evidence probing the robustness of these results. In Table W7, we interact the treatment effect with three indicators of collaborator status, to ascertain whether some among them are insulated from the effects of superstar extinction. Figure W3 provides evidence that the effect of superstar extinction decreases monotonically with the age of the collaborator at the time of death, becoming insignificantly different from zero after twenty five years of career age. Table W8 performs a number of sensitivity checks. We verify that the effect (1) is not driven by a few stars with a large number of coauthors; (2) is robust to the inclusion of indicator variables for the age of the star; (3) is not overly sensitive to our arbitrary cutoff for the superstars' age at death; and (4) is not sensitive to the problem of leakage through the coauthorship network between treated and control collaborators. Finally, we perform a small simulation study to validate the quasi-experiment exploited in the paper. We generate placebo dates of death for the control collaborators, where those dates are drawn at random from the empirical distribution of death events across years for the 112 extinct superstars. We then replicate the specification in Table III, column 1a, but we limit the estimation sample to the set of 5,064 control collaborators. Reassuringly, the effect of superstar extinction in this manufactured data is a precisely estimated zero.

2.5 Conclusion

We examine the role of collaboration in spurring the creation of new scientific knowledge. Using the premature and unexpected deaths of eminent academic life scientists as a quasi-experiment, we find that their collaborators experience a sizable and permanent decline in quality-adjusted publication output following these events. Exploiting the rich heterogeneity in these collaborative relationships, we attempt to adjudicate between plausible mechanisms that could give rise to the extinction effect.

Neither a mechanical story whereby ongoing collaborative teams struggle to replace the skills that have gone missing, nor a gatekeeping story where stars merely serve as conduits for tangible resources are sufficient to explain our results. Rather, these effects appear to be driven, at least in part, by the existence of knowledge spillovers across members of the research team. When a superstar dies, part of the scientific field to which he contributed dies along with him, perhaps because the fount of scientific knowledge from which coauthors can draw is greatly diminished. The permanence and magnitude of these effects also suggests that even collaborations which produce a small number of publications may have long-term repercussions for the pace of scientific advance.

In the end, this paper raises as many questions as it answers. It would be interesting to know whether superstar extinction also impacts the productivity of non-coauthors proximate in intellectual space, and in which direction. The degree to which exposure to superstar talent benefits industrial firms is also potentially important and represents a fruitful area that we are pursuing in ongoing research. Future work could also usefully focus on identifying quasi-experiments in intellectual space. For instance, how do scientists adjust to sudden changes in scientific opportunities in their field? Finally, collaboration incentives and opportunities may be different when scientific progress relies more heavily on capital equipment; an examination of the generalizability of our findings to other fields therefore merits further attention.

Our results shed light on an heretofore neglected causal process underlying the growth of scientific knowledge, but they should be interpreted with caution. While we measure the impact of losing a star collaborator, a full accounting of knowledge spillovers would require information on the benefits that accrued to the field while the star was alive. We can think of no experiment, natural or otherwise, that would encapsulate this counterfactual. Moreover, the benefits of exposure to star talent constitute only part of a proper welfare calculation. Scientific coauthorships also entail costs. These costs could be borne by low-status collaborators in the form of lower wages, or by the stars, who might divert some of their efforts towards mentorship activities. Though some of these costs might be offset by non-pecuniary benefits, we suspect that the spillovers documented here are not fully internalized by the scientific labor market.

Finally, for every invisible college that contracts following superstar extinction, another might expand to slowly take its place. Viewed in this light, our work does little more than provide empirical support for Max Planck's famous quip: *"science advances one funeral at a time."*

Appendix I:
Criteria for Delineating the Set of 10,349 "Superstars"

We present additional details regarding the criteria used to construct the sample of 10,349 superstars.

Highly Funded Scientists. Our first data source is the Consolidated Grant/Applicant File (CGAF) from the U.S. National Institutes of Health (NIH). This dataset records information about grants awarded to extramural researchers funded by the NIH since 1938. Using the CGAF and focusing only on direct costs associated with research grants, we compute individual cumulative totals for the decades 1977-1986, 1987-1996, and 1997-2006, deflating the earlier years by the Biomedical Research Producer Price Index.[18] We also recompute these totals excluding large center grants that usually fund groups of investigators (M01 and P01 grants). Scientists whose totals lie in the top ventile (i.e., above the 95^{th} percentile) of either distribution constitute our first group of superstars. In this group, the least well-funded investigator garnered \$10.5 million in career NIH funding, and the most well-funded \$462.6 million.[19]

Highly Cited Scientists. Despite the preeminent role of the NIH in the funding of public biomedical research, the above indicator of "superstardom" biases the sample towards scientists conducting relatively expensive research. We complement this first group with a second composed of highly cited scientists identified by the Institute for Scientific Information. A Highly Cited listing means that an individual was among the 250 most cited researchers for their published articles between 1981 and 1999, within a broad scientific field.[20]

Top Patenters. We add to these groups academic life scientists who belong in the top percentile of the patent distribution among academics — those who were granted 17 patents or more between 1976 and 2004.

Members of the National Academy of Sciences. We add to these groups academic life scientists who were elected to the National Academy of Science between 1975 and 2007.

MERIT Awardees of the NIH. Initiated in the mid-1980s, the MERIT Award program extends funding for up to 5 years (but typically 3 years) to a select number of NIH-funded investigators *"who have demonstrated superior competence, outstanding productivity during their previous research endeavors and are leaders in their field with paradigm-shifting ideas."* The specific details governing selection vary across the component institutes of the NIH, but the essential feature of the program is that only researchers holding an R01 grant in its second or later cycle are eligible. Further, the application must be scored in the top percentile in a given funding cycle.

Former and current Howard Hughes Medical Investigators. Every three years, the Howard Hughes Medical Institute selects a small cohort of mid-career biomedical scientists with the potential to revolutionize their respective subfields. Once selected, HHMIs continue to be based at their institutions, typically leading a research group of 10 to 25 students, postdoctoral associates and technicians. Their appointment is reviewed every five years, based solely on their most important contributions during the cycle.[21]

Early career prize winners. We also included winners of the Pew, Searle, Beckman, Rita Allen, and Packard scholarships for the years 1981 through 2000. Every year, these charitable foundations provide seed funding to between 20 and 40 young academic life scientists. These scholarships are the most prestigious accolades that young researchers can receive in the first two years of their careers as independent investigators.

---

[18] `http://officeofbudget.od.nih.gov/UI/GDPFromGenBudget.htm`

[19] We perform a similar exercise for scientists employed by the intramural campus of the NIH. These scientists are not eligible to receive extramural funds, but the NIH keeps records of the number of "internal projects" each intramural scientist leads. We include in the elite sample the top ventile of intramural scientists according to this metric.

[20] The relevant scientific fields in the life sciences are microbiology, biochemistry, psychiatry/psychology, neuroscience, molecular biology & genetics, immunology, pharmacology, and clinical medicine.

[21] See Azoulay et al. (2009) for more details and an evaluation of this program.

Table 2.1
Summary Statistics for Superstar Scientists (N=112)

Sample consists of 112 superstar life scientists who died suddenly while still actively engaged in research. See Appendix I and Section II.A for details on sample construction. Degree year denotes the year of the most recent degree attained by the superstar. Number of collaborators is defined as the number of distinct coauthors within the scientists' cumulative stock of publications. NIH review panel membership denotes stars who were members of an NIH review panel in the five years prior to their death, and the number of collaborators in NIH review panels refers to the number of coauthors of each superstar who sat on NIH review panels in the 5 years prior to the star's death. We use the terms "star" and "superstar" interchangeably.

	Mean	Median	Std. Dev.	Min.	Max.
Birth Age at Death	57.17	58	7.042	37	67
Degree year	1962.741	1964	10.193	1942	1984
MD	0.42	0	0.496	0	1
PhD	0.438	0	0.498	0	1
MD/PhD	0.143	0	0.351	0	1
Female	0.063	0	0.243	0	1
U.S. Born	0.786	1	0.412	0	1
Nb. of Collaborators	47.027	37	34.716	3	178
NIH Review Panel Membership (past 5 yrs)	0.045	0	0.207	0	1
Nb. of Collabs. in NIH Review Panels (past 5 yrs)	1.33	1	1.657	0	7
Career Nb. of Publications	139.607	121	91.371	25	473
Career Nb. of Citations	8,190	6,408	7,593	435	38,941
Career NIH Funding	$10,722,590	$8,139,397	$12,057,638	$0	$70,231,584

Figure 2-1
Distribution of Coauthorship Intensity

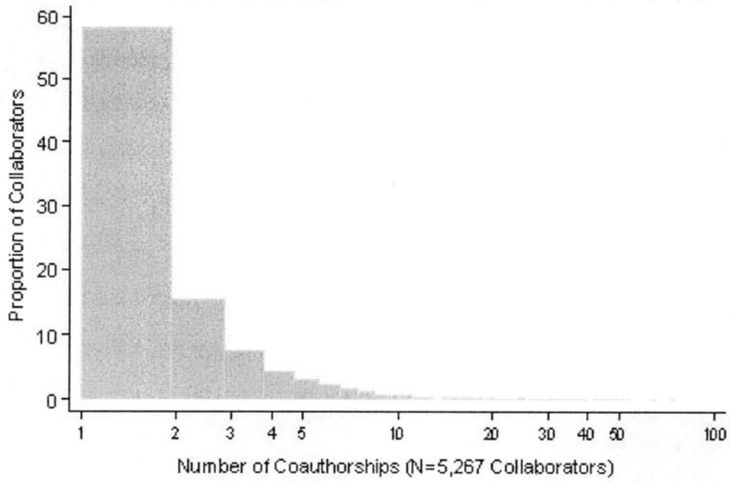


Table 2.2
Summary Statistics for Collaborators in the Year of Superstar Death

The samples consist of faculty collaborators of 112 extinct superstar life scientists an equal number of matched control coauthors. See Sections II.B and III for details on the sample construction and variable definitions and Appendix II for details on the matching procedure. All variables are measured as of the year of superstar death. Publications are JIF-weighted.

		Mean	Median	Std. Dev.	Min.	Max.
Control	Nb. of weighted Publications	18.314	8	27.917	0	342
Collabs.	Cum. Nb. of weighted Publications	327.33	187	409.098	0	3,968
(N=5,064)	Holds R01 grant	0.559	1	0.497	0	1
	Co-Located	0.144	0	0.351	0	1
	Career Age	23.698	23	9.963	1	59
	Elite	0.093	0	0.29	0	1
	Cum. Nb. of Coauthorships	2.734	1	4.339	1	69
	Nb. of Other Superstar Collabs.	2.746	2	3.516	0	31
	Years since first Coauthorship	10.949	10	7.901	0	42
	Years since last Coauthorship	9.275	8	7.774	0	41
	Former trainee of the star	0.07	0	0.255	0	1
	"Accidental" Collaborator	0.076	0	0.265	0	1
	MeSH Keyword Overlap	0.265	0	0.162	0	1
	Superstar Citation Count	10,083	7,245	8,878	99	90,136
Treated	Nb. of weighted Publications	19.068	8	31.656	0	491
Collabs.	Cum. Nb. of weighted Publications	334.905	187	436.927	0	4,519
(N=5,064)	Holds R01 grant	0.571	1	0.495	0	1
	Co-Located	0.123	0	0.328	0	1
	Career Age	23.761	23	9.969	0	59
	Elite	0.077	0	0.266	0	1
	Cum. Nb. of Coauthorships	2.835	1	4.894	1	75
	Nb. of Other Superstar Collabs.	3.087	2	4.255	0	44
	Years since first Coauthorship	11.022	10	7.896	0	39
	Years since last Coauthorship	9.255	8	7.728	0	38
	Former trainee of the star	0.084	0	0.278	0	1
	"Accidental" Collaborator	0.075	0	0.264	0	1
	MeSH Keyword Overlap	0.259	0	0.157	0	1
	Superstar Citation Count	10,228	7,239	7,952	397	34,746

Table 2.3
Impact of Superstar Death on Collaborators' Publication Rates

Estimates stem from conditional quasi-maximum likelihood Poisson specifications. Dependent variable is the total number of JIF-weighted articles authored by a collaborator of a superstar life scientist in the year of observation. All models incorporate a full suite of year effects as well as 17 age category indicator variables (career age less than -3 is the omitted category). Exponentiating the coefficients and differencing from one yield numbers interpretable as elasticities. For example. the estimates in column (1a) imply that collaborators suffer on average a statistically significant (1-exp[-0.092])=8.79% decrease in the rate of publication after their superstar coauthor passes away. Robust (QML) standard errors in parentheses, clustered at the level of the superstar. $p < 0.10$, *$p < 0.05$, **$p < 0.01$

	Panel A All JIF-Weighted Publications		Panel B JIF-Weighted Pubs. Written with others	
	Without Ctrls (1a)	With Ctrls (1b)	Without Ctrls (2a)	With Ctrls (2b)
After Death	−0.092 ∗ ∗	−0.086 ∗ ∗	−0.057 ∗ ∗	−0.054∗
	(0.022)	(0.025)	(0.022)	(0.024)
Log Pseudo-Likelihood	-974,285	-1,832,594	-950,864	-1,783,958
Nb. of Obs.	153,508	294,943	153,508	294,943
Nb. of Collaborators	5,267	10,128	5,267	10,128

Figure 2-2
Proximity in Ideas Space

Measure of distance in ideas space is defined as the number of unique MeSH terms which overlap between the colleague's and superstar's publications (excluding coauthored output), normalized by the total number of MeSH terms used in the colleague's total publications. This measure is calculated for articles published in the five years preceding superstar death.

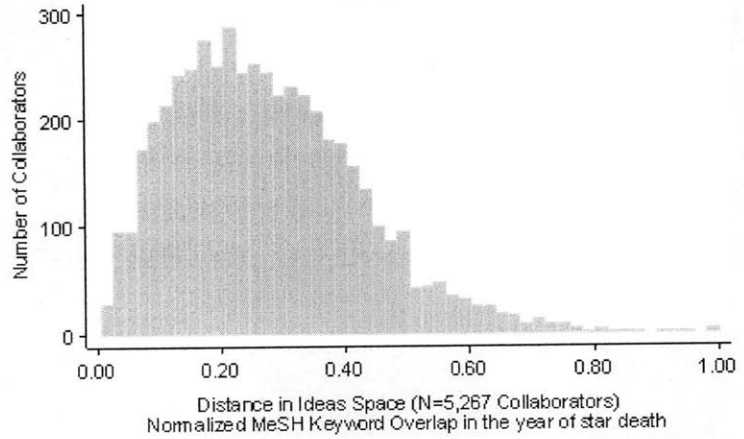


Table 2.4
Collaborator Publication Rates and Imperfect Skill Substitution

Estimates stem from conditional quasi-maximum likelihood Poisson specifications. Dependent variable is the total number of JIF-weighted articles authored by a collaborator of a superstar life scientist in the year of observation. Regular and Close Collaborator are indicator variables for the number of publications coauthored by the superstar and colleague at the time of death (regular collaborations correspond to between 3 and 9 coauthored pubs.; close collaborations correspond to 10 or more coauthored pubs; casual collaborations - the omitted category - corresponds to 1 or 2 coauthored pubs.). All models incorporate year effects and 17 age category indicator variables (career age less than -3 is the omitted category), as well as 17 interaction terms between the age effects and each covariate of interest (i.e., column (3b) includes a total of 3X17=51 age-specific interaction terms). Robust (QML) standard errors in parentheses, clustered at the level of the superstar. $p < 0.10$, *$p < 0.05$, **$p < 0.01$

	Coauthorship Intensity		Coauthorship Recency		Coauthorship Intensity & Recency	
	All Pubs.	Pubs. written with others	All Pubs.	Pubs. written with others	All Pubs.	Pubs. written with others
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
After Death	−0.076**	−0.057*	−0.087**	−0.074**	−0.080**	−0.075**
	(0.026)	(0.025)	(0.024)	(0.024)	(0.024)	(0.024)
After Death X Regular Collab.	−0.044	−0.02			−0.039	−0.018
	(0.041)	(0.042)			(0.042)	(0.043)
After Death X Close Collab.	−0.026	0.117			−0.014	0.119
	(0.068)	(0.073)			(0.069)	(0.074)
After Death X At least 1 coauthorship in the three years preceding stars death			−0.022	0.032	−0.021	0.028
			(0.038)	(0.039)	(0.039)	(0.039)
Log Pseudo-Likelihood	-1,831,987	-1,781,742	-1,822,664	-1,775,680	-1,821,791	-1,774,167
Nb. of Obs.	294,943	294,943	294,943	294,943	294,943	294,943
Nb. of Collabs.	10,128	10,128	10,128	10,128	10,128	10,128

Table 2.5
Collaborator Publication Rates and Access to Resources

Estimates stem from conditional quasi-maximum likelihood Poisson specifications. Dependent variable is the total number of JIF-weighted articles authored by a collaborator of a superstar life scientist in the year of observation. Betweenness centrality is measured using the network of 10,349 superstar life scientists, Former trainee indicates that the colleague was a graduate student or postdoctoral fellow in the laboratory of the superstar (7.69% of the collaborators). All models incorporate year effects and 17 age category indicator variables (career age less than -3 is the omitted category), as well as 17 interaction terms between the age effects and each covariate of interest. Robust (QML) standard errors in parentheses, clustered at the level of the superstar.

	Star's Ties to NIH Funding Process (1)	Quartile of Betweenness Centrality (2)	Former Trainee (3)	All Covariates Combined (4)
After Death	−0.105**	−0.067*	−0.086**	−0.089*
	(0.037)	(0.028)	(0.025)	(0.035)
After Death X Star Sat	0.042			0.024
on NIH Review Panel	(0.064)			(0.07)
After Death X Stars Nb.	0.011			0.014
of Coauth. Ties to NIH Review Panelists	(0.013)			(0.015)
After Death X Star in		−0.031		−0.04
4th Quartile of Betweenness Centrality		(0.046)		(0.051)
After Death X Coauthor is			0.056	0.048
Former Trainee			(0.069)	(0.069)
Log Pseudo-Likelihood	-1,831,339	-1,831,779	-1,830,582	-1,828,754
Nb. of Obs.	294,943	294,943	294,943	294,943
Nb. of Collabs.	10,128	10,128	10,128	10,128

Table 2.6
Collaborator Publication Rates and Proximity in Geographic & Intellectual Space

Estimates stem from conditional quasi-maximum likelihood Poisson specifications. Dependent variable is the total number of JIF-weighted articles authored by a collaborator of a superstar life scientist in the year of observation. Co-located indicates that the colleague and superstar were employed at the same institution at the time of superstar death. Keyword overlap is the normalized number of MeSH keywords which appear on both the colleague and superstars non-joint publications. Accidental collaborators are those who only appear on coauthored publications with the superstar when both are in the middle of the authorship list. Regular and Close Collaborator are indicator variables for the number of publications coauthored by the superstar and colleague at the time of death (regular collaborations correspond to between 3 and 9 coauthored pubs.; close collaborations correspond to 10 or more coauthored pubs; casual collaborations - the omitted category - corresponds to 1 or 2 coauthored pubs.). All models incorporate year effects and 17 age category indicator variables (career age less than -3 is the omitted category), as well as 17 interaction terms between the age effects and the covariate of interest. Robust (QML) standard errors in parentheses, clustered at the level of the superstar.

	(1)	(2)	(3)	(4)	(5)
After Death	−0.092**	−0.067**	−0.094**	−0.081**	−0.074**
	(0.027)	(0.023)	(0.022)	(0.024)	(0.026)
After Death X Co-Located	0.042			0.037	0.042
	(0.043)			(0.043)	(0.044)
After Death X Kwd. Overlap in Top Quartile		−0.115*		−0.114	−0.127*
		(0.059)		(0.059)	(0.057)
After Death X "Accidental" Collaborator			0.104	0.111	0.077
			(0.06)	(0.058)	(0.055)
After Death X Regular Collaborator					−0.03
					(0.044)
After Death X Close Collaborator					0.002
					(0.072)
After Death X Recent Collaborator					−0.022
					(0.038)
% of Collabs. Affected	13.33	25.35	7.53		
Log Pseudo-Likelihood	-1,831,900	-1,830,305	-1,831,787	-1,828,805	-1,817,667
Nb. of Obs.	294,943	294,943	294,943	294,943	294,943
Nb. of Collabs.	10,128	10,128	10,128	10,128	10,128

Table 2.7
Impact of Superstar Status on Collaborators' Publication Rates

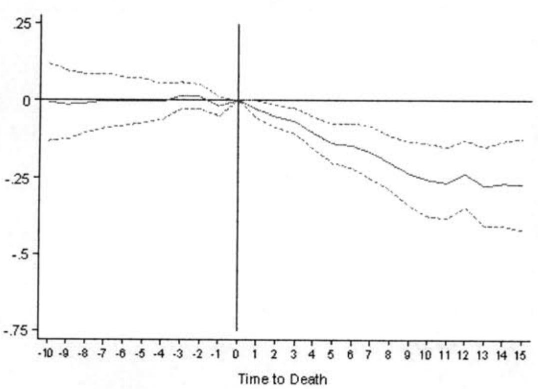
Estimates stem from conditional quasi-maximum likelihood Poisson specifications. Dependent variable is the total number of JIF-weighted articles authored by a collaborator of a superstar life scientist in the year of observation. Top quartiles of citations and career NIH funding are defined using the population of 10,009 superstar scientists with appointments compatible with extramural NIH funding. We exclude from the estimation sample the collaborators of 11 "intramural" NIH scientists who are not eligible to receive extramural funding. All models incorporate year effects and 17 age category indicator variables (career age less than -3 is the omitted category), as well as 17 (columns (1) and (2)) or 34 (column (3)) interaction terms between the age effects and the "Top Quartile" indicator variable. Robust (QML) standard errors in parentheses, clustered at the level of the superstar.

	Superstar Status Citations	Superstar Status NIH Funding	Superstar Status Citations & NIH Funding
	(1)	(2)	(3)
After Death	−0.034	−0.070*	−0.026
	(0.036)	(0.035)	(0.039)
After Death X Star in Top Quartile of Cites	−0.082		−0.080
	(0.047)		(0.048)
After Death X Star in Top Quartile of NIH $		−0.026	−0.016
		(0.05)	(0.051)
Log Pseudo-Likelihood	-1,715,929	-1,716,213	-1,715,916
Nb. of Obs.	275,776	275,776	275,776
Nb. of Collabs.	9,470	9,470	9,470

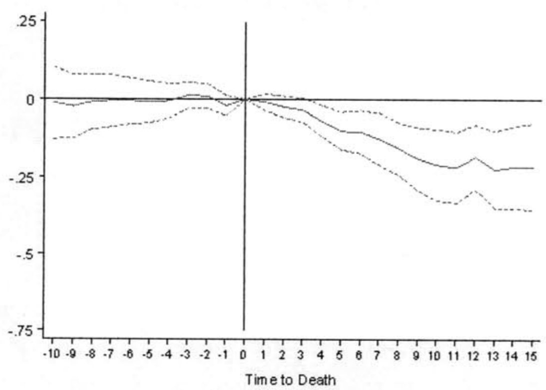
Figure 2-3
Dynamics of the Treatment Effect

The solid blue lines in the above plots correspond to coefficient estimates of conditional fixed effects quasi-maximum likelihood Poisson specifications in which the weighted publication output of a collaborator is regressed onto year effects, 17 indicator variables corresponding to different age brackets, and interactions of the treatment effect with 27 indicator variables corresponding to 11 years before the year of death and prior, 10 years before the year of death, 9 years before the year of death,, 14 years after the year of death, and 15 years after the year of death and above (the indicator variable for treatment status interacted with the year of death is omitted). The 95% confidence interval (corresponding to robust standard errors, clustered around superstars) around these estimates is plotted with dashed red lines. Figure IIIA uses column (1b) of Table III as a baseline (i.e., treated and control collaborators, the dep. var. includes all of the collaborator's publications); Figure IIIB uses column (2b) of Table III as a baseline (i.e., treated and control collaborators, the dep. var. is limited to the collaborator's publications in which the superstar does not appear on the authorship list).

A: All publications



B: Publications without Superstar Collaborator



Appendix II: Construction of the Control Group

We detail the procedure implemented to identify the control collaborators that help pin down the life-cycle and secular time effects in our difference-in difference (DD) specification. Because it did not prove possible to perfectly match treated and control collaborators on all covariates, the procedure is guided by the need to guard against two specific threats to identification.

First, collaborators observed in periods before the death of their associated superstar are more likely to work with a younger superstar; thus, they are not ideal as a control if research trends of collaborators differ by the age of the superstar. Collaborators observed in periods after the death of their associated superstar are only appropriate controls if the death of their superstar only affected the level of their output; if the death also negatively affected the trend, fixed effects estimates will be biased towards zero.

Second, fixed effects estimates might be misleading if collaborations with superstars are subject to idiosyncratic dynamic patterns. Happenstance might yield a sample of stars clustered in decaying scientific fields. More plausibly, collaborations might be subject to specific life-cycle patterns, with their productive potential first increasing over time, eventually peaking, and thereafter slowly declining. Relying solely on collaborators treated earlier or later as as an implicit control group entails that this dyad-specific, time-varying omitted variable will not be fully captured by collaborator age controls.

To address these threats, the sample of control collaborators (to be recruited from the universe of collaborators for the 10,000 stars who do not die prematurely, regardless of cause) should be constructed such that the following four conditions are met:

1. treated collaborators exhibit no differential output trends relative to control collaborators up to the time of superstar death;
2. the distributions of career age at the time of death are similar for treated and controls;
3. the time paths of output for treated and controls are similar up to the time of death;
4. the dynamics of collaboration up to the time of death — number of coauthorships, time elapsed since first/last coauthorship, superstar's scientific standing as proxied by his cumulative citation count — are similar for treated and controls.

Coarsened Exact Matching. To meet these goals, we have implemented a "Coarsened Exact Matching" (CEM) procedure (Iacus et al. 2008) to identify a control for each treated collaborator. As opposed to methods that rely on the estimation of a propensity score, CEM is a non-parametric procedure. This seems appropriate given that we observe no covariates that predict the risk of being associated with a superstar scientist who dies in a particular year.

The first step is to select a relatively small set of covariates on which the analyst wants to guarantee balance. In our example, this choice entails judgement, but is strongly guided by the threats to identification mentioned above. The second step is to create a large number of strata to cover the entire support of the joint distribution of the covariates selected in the previous step. In a third step, each observation is allocated to a unique strata, and for each observation in the treated group, a control observation is selected from the same strata; if there are multiple choices possible, ties are broken randomly.

The procedure is coarse because we do not attempt to precisely match on covariate values; rather, we coarsen the support of the joint distribution of the covariates into a finite number of strata, and we match a treated observation if and only if a control observation can be recruited from this strata. An important advantage of CEM is that the analyst can guarantee the degree of covariate balance *ex ante*, but this comes at a cost: the more fine-grained the partition of the support for the joint distribution (i.e., the higher the number of strata), the larger the number of unmatched treated observations.

Implementation. We identify controls based on the following set of covariates (t denotes the year of death): collaborator's degree year, number of coauthorships with the star at t, number of years elapsed since last coauthorship with the star at t, JIF-weighted publication flow in year t, cumulative stock of JIF-weighted publications up to year $t-1$, and the star's cumulative citation count at t. We then coarsen the joint distributions of these covariates by creating 51,200 strata. The distribution of degree years is coarsened using three year intervals; the distribution of coauthorship intensity is coarsened to map into our taxonomy of casual (1 and 2 coauth.), regular (between 3 and 9 coauth.), and close collaborators (10 or more coauth.); the distribution of coauthorship recency is coarsened into quartiles (the first quartile corresponds to recent relationships, i.e. less than 3 years since the last coauthorship); the flow of publications in the year of death is coarsened into 5 strata (the three bottom quartiles; from the 75^{th} to the 95^{th} percentile, and above the 95^{th} percentile); the stock of publications is coarsened into eleven strata (0 to 5^{th}; 5^{th} to 10^{th}; 10^{th} to 25^{th}; 25^{th} to 35^{th}; 35^{th} to 50^{th}; 50^{th} to 65^{th}; 65^{th} to 75^{th}; 75^{th} to 90^{th}; 90^{th} to 95^{th}; 95^{th} to 99^{th}; and above the 99^{th} percentile); and the distribution of citation count for the star is coarsened into quartiles.
We implement the CEM procedure year by year, without replacement. Specifically, in year t, we:

1. eliminate from the set of potential controls all superstars who die, all coauthors of superstars who die, and all control coauthors identified for years of death k, $1979 \leq k < t$;

2. create the strata (the values for the cutoff points will vary from year to year for the some of the covariates mentioned above);

3. identify within strata a control for each treated unit; break ties at random;

4. repeat these steps for year $t+1$.

We match 5,064 out of 5,267 treated collaborators (96.15%). In the sample of $5,064$ treated$+5,064$ controls$=10,128$ collaborators, there is indeed no evidence of prexisting trends in output (Figure A1); nor is there evidence of differential age effects in the years leading up to the death event (Figure A2). As seen in Table II, treated and controls are very well-balanced on all covariates that pertain to the dynamics of the collaboration: number of coauthorships, time since last and first coauthored publication, and superstar's citation count. The age distributions are very similar as well. Furthermore, the CEM procedure balances a number of covariates that were not used as inputs, such as normalized keyword overlap and R01 NIH grantee status. For some covariates, we can detect statistically significant mean differences, though they do not appear to be substantively meaningful (e.g., 7% of controls vs. 8.4% of treated collaborators were former trainees of their associated superstars).

Sensitivity Analyses. The analyst's judgement matters for the outcome of the CEM procedure insofar as she must draw a list of "reasonable" covariates to match on, as well as decide on the degree of coarsening to impose. Therefore, it is reasonable to ask whether seemingly small changes in the details have consequences for how one should interpret our results.
Non-parametric matching procedures such as CEM are prone to a version of the "curse of dimensionality" whereby the proportion of matched units decreases rapidly with the number of strata. For instance, requiring a match on an additional indicator variable (i.e., doubling the number of strata from around 50,000 to 100,000) results in a match rate of about 70%, which seems unacceptably low. Conversely, focusing solely on degree age and the flow and stock of the outcome variables would enable us to achieve pairwise balance (as opposed to global balance, which ignores the one-to-one nature of the matching procedure) on this narrower set of covariates, but at the cost of large differences

in the features of collaborations (such as recency and intensity) between treated and controls. This would result in a control sample that could address the first threat to identification mentioned above, but not the second.
However, we have verified that slight variations in the details of the implementation (e.g., varying slightly the number of cutoff points for the stock of publications; focusing on collaboration age as opposed to collaboration recency; or matching on superstar funding as opposed to superstar citations) have little impact on the results presented in Table III.
To conclude, we feel that CEM enables us to identify a population of control collaborators appropriate to guard against the specific threats to identification mentioned in section 2.3.

Appendix III: Matching Superstars and Collaborators

We designed the Stars/Colleague Generator (S/CGEN) to harvest coauthors' names from a superstar's bibliome. S/CGEN identifies colleagues to the extent that (a) they coauthor at least once; and (b) they can be matched (based on a combination of a last name and up to two initials) with the AAMC Faculty Roster. We will describe the matching process using as an example one of our extinct superstars, Jeffrey M. Isner, MD. Isner, a pioneer of gene therapy for Peripheral Artery Diseases, and a faculty member at the Tufts University School of Medicine, died in 2001 from a heart attack, at the age of 54.

The matching process begins with the creation of a customized PubMED search query for each superstar. In the case of Isner, the query is `("isner jm"[au] OR "isner j"[au]) AND 1977:2006[dp]`, and it returns 373 original publications (the query also returns 24 letters, editorials, interviews, etc., which we ignore). The process of harvesting bibliomes from PubMED using name variations and queries as inputs is facilitated by the use of PUBHARVESTER, a software program we specifically designed for this purpose (Azoulay et al. 2006).

Spurious Coauthors. Jeff Isner's PubMED query accounts for his inconsistent use of the middle initial, but is otherwise quite simple. For other scientists, queries might factor in their inconsistent use of the suffix "Jr.," or name variations coincident with changes in marital status. For yet many others with frequent names, the queries are more involved, and make use of CV information such as scientific keywords, institutional affiliation, frequent coauthors' names, etc. This is essential, since errors of commission will tend to generate spurious coauthor matches. We guarded against this source of error by devoting hundreds of person-hours to the design of accurate search queries for each of our 10,349 superstars. This degree of labor-intensive customization ensures that a superstar's bibliome excludes publications belonging to homonymous scientists.

Matching process. The second step is to extract the name of coauthors from the star's bibliome and to match them with the AAMC Faculty Roster. Unfortunately, PubMED does not record authors' full names, nor does it record their institutional affiliations; it only keeps track of authors by using a combination of last name, two initials, and a suffix (where the suffix and the second initial fields can be empty). The matching process is automated by SC/GEN, and its outcome in the case of a sample publication authored by Jeff Isner is illustrated in Figure W1. S/CGEN cannot generate a match for each coauthor. Some coauthors are technicians or undergraduate students; others are graduate students or postdocs who do not go on to faculty positions; yet others are located in foreign institutions; others still publish under names that differ from the Faculty Roster listing (for instance by being inconsistent with the use of middle initials, suffixes, or hyphens). In total, SC/GEN generates 355 matches with the AAMC Faculty Roster for Isner.

Ambiguous Coauthors. Often, SC/GEN can match a given PubMED name with more than one faculty in the Roster. Notice the case of `ramaswamy k` on Figure W1. Does it correspond to K. Ramaswamy (University of Illinois–Chicago), to Karthik Ramaswamy (UMASS School of Medicine), or to Krishna Ramaswamy (Tufts University School of Medicine)? Several options are available to deal with these ambiguous matches. We could discard the first two matches, since the third one corresponds to an individual who shared Isner's institutional affiliation. Alternatively, we could retain all three matches, but assign each a weight of $\frac{1}{3}$, incorporating a guess on the probability that each match is genuine. Finally, we could simply discard all three matches, and focus instead on those matches that are unambiguous. This is the approach we have followed to generate the results we present in the paper.[22] Out of the

[22] Trajtenberg et al. (2006) propose algorithms to automate the process of name disambiguation in patent data. Adapting their approach to publication data lies far beyond the scope of this paper. To fix ideas, `Lechleiter JD` is an example of unique PubMED name. In contrast, `Weinstein SL` corresponds to two distinct faculty in the roster, `Miller MJ` to ten, and `Wang Y` to thirty six.

355 matches mentioned above, only 177 correspond to coauthors with unambiguous PubMED names. For the set of 112 superstars, S/CGEN identifies 5,267 distinct coauthors with unambiguous PubMED names — an average of 47 coauthors per superstar (the median is 37).

Coauthors' Publication Output. The publication output of coauthors with frequent names will be measured with error. This source of error is less worrisome, since it involves a dependent variable. Nonetheless, we have taken several steps to ascertain the extent to which it biases our results. First, our decision to eliminate from the sample coauthors with ambiguous PubMED names means that it is almost entirely composed of individuals with relatively rare names. Second, we have experimented with deleting from the estimation sample observations corresponding to coauthors with unique PubMED names, but popular last names.[23] Specifically, we dropped from the main analysis all coauthors whose last name appear 160 or more times in the roster (the 99^{th} percentile of the distribution of last name frequency, which correspond to names such as `Greenwald`, `McKee`, `O'Malley`, or `Fu`). This hardly affected the main results. Third, we limit the estimation sample to elite coauthors (i.e., coauthors who belong to the set of 10,349 "superstars"). Because we designed custom PubMED queries for these individuals, their output is measured with little (if any) error. The magnitude of the treatment effect is very similar to the one obtained on the full sample of coauthors.

Appendix IV: Measuring Proximity in Ideas Space

We describe the construction of our variable to measure distance (or rather, proximity) in intellectual or "ideas space" between nodes in a dyad of scientists. The boundaries around scientific fields are difficult to delineate since most scientific research can be classified in numerous ways, and agreement among scientists regarding the categorization of specific bits of knowledge is often elusive. Our approach is predicated on the inadequacy of measures based on shared department affiliation, or on coarse distinctions between scientific fields (e.g., cell vs. molecular biology). Instead of attempting to position individual scientists relative to some fixed address in ideas space, we provide a method to cheaply and conveniently measure *relative position* in this space.

An essential input is provided by the Medical Subject Headings (MeSH) thesaurus, a controlled vocabulary produced by the National Library of Medicine whose explicit statement of purpose is to *"provide a reproducible partition of concepts relevant to biomedicine for the purpose of organizing knowledge and information."* The MeSH vocabulary consists of 24,767 terms arranged in a hierarchical structure, and these terms are used by NLM staff to tag all the articles indexed by the PubMED database.[24] From our standpoint, one of the MeSH system's most attractive feature is its fine-grained level of detail. For instance, the initial draft of the public human genome project (Lander et al. 2001) is tagged by 26 distinct descriptors, which run the gamut from the very general ("Humans", "RNA/Genetics") to the very specific ("Repetitive Sequences, Nucleic Acid", "CpG Islands", "DNA Transposable Elements").[25]

The procedure followed to generate our dyadic measure of intellectual proximity is best explained through a concrete example. We will focus on a two scientists, Andrew Schally (from Tulane University in New Orleans, LA) and Roger

[23] For instance, `Miller CR` is a unique PubMED name, though `Miller` is the last name for 800 distinct individuals in the AAMC Faculty Roster.

[24] At the highest level of the hierarchical structure are very broad headings such as "Anatomy" or "Mental Disorders." More specific headings are found at lower levels of the eleven-level hierarchy, such as "Ankle" and "Conduct Disorder." See `http://www.nlm.nih.gov/mesh/` for more details.

[25] This stands in sharp contrast to the coarse partition of technological space provided by patent classes, which are often used in the study of *involuntary* knowledge spillovers (Benner and Waldfogel 2007).

Guillemin (from the Salk Institute in San Diego, CA). Throughout the 1960s and 1970s, this pair of eminent neuro-endocrinologists was locked in a very public (and often acrimonious) rivalry whose ultimate goal was the synthesis of peptide hormones produced by the brain. Together with Rosalyn Yalow, the Nobel committee awarded them both the Prize in Medicine and Physiology in 1977 (details of this celebrated case of a scientific race can be found in Nicholas Wade's *The Nobel Duel*). We will focus on the five-year window that preceded the award of the Prize, i.e., 1973-1977. During this period, Guillemin and Schally did not collaborate at all, and according to Wade (1981), even actively sought to undermine each other's progress.

The calculation is illustrated in Table W2; it is automated by SCIDIST, an open-source software program we specifically designed for this purpose.[26] Between 1973 and 1977, Schally published 240 articles, and Guillemin "only" 60. We extract from these publications all MeSH terms, regardless of their position in the descriptor hierarchy. There are a total of 607 unique MeSH terms tagging the two scientists' publications, 147 of which overlap. Table W2 lists the Top 10 overlapping terms with highest and lowest combined use, respectively.[27]

To compute the proximity of Guillemin to Schally, we simply divide the number of overlapping MeSH terms (147), by the total number of unique MeSH terms tagging Guillemin's 60 publications (220). In contrast, the proximity of Schally to Guillemin is given by 147 divided by 534 (the total number of unique MeSH terms tagging Schally's 240 publications). We view this lack of symmetry as an attractive feature of our approach, since Schally's research agenda during this period was significantly broader, and in fact encompassed most of Guillemin's. In contrast, many of the distance concepts used to date in the literature — for example to position firms' research portfolio in technology space — use an Euclidean (hence symmetric) concept of distance (e.g., Jaffe 1986).

Appendix V: Pre-existing Trends in Output for the Superstars

In Table W3, we present results for specifications in which the superstars' quality-adjusted publication output is regressed onto a series of indicator variables corresponding to the timing of death: 5 years before the year of death, 4 years before the year of death, and so on, up until two years after the year of death (a scientist can, and often does, publish after his death because his/her coauthors will typically steward articles through the pipeline on his behalf). All models include superstar scientist fixed effects, and we use as a control group the set of superstars who collaborate with the sample of control collaborators. The inclusion of controls is important insofar as it enables us to pin down the effect of age and calendar time, which might be correlated with the death effect.

We use two definitions of the dependent variable. In the first (column 1), all of the stars' publications participate in the calculation of the JIF-weighted count; in the second (column 2), only the publications in which the star appears in last position on the authorship roster are considered (last author status is reserved to the heads of laboratory/research group in the life sciences). In both specifications, we find no evidence that the superstars' output trends downward even before their death. In fact, the coefficient estimates turn negative in sign only in the year that follows the year of death, and reach statistical significance only two years after the death. In light of these results, we feel confident that our informal screen for research activity yields a set of 112 extinct superstars still actively engaged in science at the time of their deaths.

[26] SCIDIST is available for download at `http://www.stellman-greene.com/ScientificDistance/`.

[27] An open question is whether one should weight each term by its frequency of use, or whether it is the number of unique terms that matters. In practice, these alternatives yield two measures of proximity that are heavily correlated, and the distinction does not affect the substance of our results.

Appendix VI: Main Results with OLS Estimation

In Table W4, we replicate the results in Table III using linear collaborator fixed effects specifications. This robustness check is informative insofar as linear specifications enable us to completely saturate the specifications with age effects (a total of 54 indicator variables, vs. 17 in the QML Poisson specifications presented in the main body of the paper). The results in column 1a imply that coauthors suffer a 1.55 yearly decline in JIF-weighted publication output following the death of their superstar collaborator. This represents a 8.14% decrease relative to the mean of the dependent variable at the time of death. In contrast, the estimate of the treatment effect in column 1a of Table III corresponds to a 8.79% decline in the JIF-weighed publication rate. The magnitudes observed in colums 1b through 2b in Table III and W4 are likewise very similar.

Appendix VII: Publication-level Quality Adjustment using Citation Data

The quality adjustment used to produce JIF-weighted publication counts is crude. It does not allow us to learn whether the research that does not get published as a consequence of superstar death is more likely to be of great vs. marginal significance. Table W5 answers this question by modeling the effect of superstar extinction for the production of articles falling above various quantiles of the citation distribution. An important caveat is that the results pertain only to the set of 1,436 controls$+1,416$ treated$=2,852$ collaborators who are also part of our elite group of 10,349 scientists, since this is the set for which article-level citation data is available. These 2,852 scientists account for 28.16% of the collaborators in the overall sample.

Citation data suffer from a well known truncation problem: older articles have had more time to be cited, and hence are more likely to reach the tail of the citation distribution, ceteris paribus. To overcome this issue, we compute a different empirical cumulative distribution for the article-level distribution of citations *in each publication year.*[28] For example, in the life sciences broadly defined, an article published in 1980 would require at least 98 citations to fall into the top ventile of the distribution; an article published in 1990, 94 citations; and an article published in 2000, only 57 citations (this is illustrated in Figure W2). With these empirical distributions in hand, it becomes meaningful to count the number of articles that fall, for example, in the top quartile of citations for a given scientist in a particular year. These counts in turn provide the dependent variables used in Table W5.

We begin by replicating the results of Table III, Model 2b on this restricted set of collaborators. The treatment effect is slightly lower in magnitude, but remains highly statistically significant (column 1). The same result obtains when using the raw (i.e., not JIF-weighted) number of publications as the outcome variable (column 2). We then find that the magnitude of the treatment effect increases as one restricts the dependent variable to publications that fall in higher quantiles of the citation distribution. It hovers between -6 and -9% when we examine the effect of superstar extinction on publications that fall in the bottom quartile, below the median, above the median, or in the top quartile of the citation distribution. It increases to -9% for publications in the top ventile, and still further to -15% when focusing on "blockbuster" publications — those falling in the top percentile of the citation distribution. At the very least, these results suggest that superstar exposure is not limited to the production of relatively less significant scientific knowledge.

[28] We thank Stefan Wuchty and Ben Jones from Northwestern University for performing the computations. These vintage-specific distributions are not based on in-sample article data, but use the universe of articles published since 1970 in biomedical and chemical journals indexed by the Web of Science.

Appendix VIII: Effect of Superstar Extinction on Receipt of NIH Funding

We present evidence on the effect of superstar extinction on receipt of NIH funding. Grants are typically awarded for a period of years (three to five is typical), and disbursed in equal yearly amounts over this period. Only the first of these payments is indicative of successful grantsmanship. We exclude from the calculation non-research grants (fellowships, training grants, and infrastructure grants), as well as large center research grants. The CGAF dataset only lists principal investigators (PIs) for each grant; as a result, we are unable to separate the grants in which coauthor and superstars are co-investigators from those that do not entail a formal research collaboration. This limitation must be borne in mind when interpreting the results of specifications relying on grant data. We also eliminate from the estimation sample 318 treated and 260 control collaborators who are NIH employees at some point during their career, and as such not eligible for receipt of extramural NIH funding.

Table W6 presents the results, using two different dependent variables: the number of research grants (columns 1a and 1b), and the probability of receiving at least one grant in a given year (columns 2a and 2b). The first two models are estimated using conditional collaborator fixed effects quasi-maximum likelihood. In these specifications, the 3,669 collaborators (38.72% of the controls vs. 36.55% of the treated) who never receive a grant during the observation period drop out of the observation sample. The last two models are estimated using a collaborator fixed effects linear probability model, on the entire sample of grant-eligible collaborators, including 37% among them who never receive (and may not even have applied for) a grant from the NIH.

The magnitudes of the effects in columns 1a and 1b are strikingly similar to those observed for publication output, though they are only statistically significant at the 10% level. In contrast, the magnitudes of the extinction effect for the linear probability models are quite small: they suggest that the probability of receiving a grant falls by a statistically significant 1% after the scientist loses a superstar collaborator. We must interpret these results with caution: there is obviously large heterogeneity in the quality and importance of research grants, and our dependent variable does not account for this.

Appendix IX: Treatment Effect Heterogeneity: Impact of Collaborator Status and Age at the Time of Superstar Death

In Table W7, we interact the treatment effect with three indicators of collaborator status, to ascertain whether some among them are insulated from the effects of superstar extinction documented earlier. Column 1 focuses on faculty members whose sole elite collaborator was the superstar who died. For these coauthors with relatively poor substitution opportunities (they account for roughly 27.66% of the dyads in the sample), the consequences of the superstar's loss are particularly severe, with an overall 15.3% decline in publication output. Columns 2 asks whether scientists who are PIs on a NIH R01 grant at the time of their superstar coauthor's death are shielded from the adverse effects documented earlier. With independent funding of this type, these investigators (who account for more than half of the sample) are likely to be less dependent on the goodwill of their collaborators, but we find no evidence supporting this conjecture: the differential effect is an small and imprecisely estimated; Independent NIH funding is not enough to insulate scientists from the loss of an eminent collaborator. In column 3, we present evidence that the "elite among the elite" (members of the National Academy of Science, Howard Hughes Medical Investigators, and NIH MERIT awardees who together account for 8.5% of the total number of collaborators) is relatively unaffected by the loss of a "peer superstar." The differential impact on elite coauthors is positive, large, and statistically significant; it offsets almost exactly the main treatment effect.

We conclude that the effect of superstar extinction is heterogeneous with respect to coauthor status, but the heterogeneity stems from the tails of the status distribution. The loss of a prominent collaborator adversely impacts the productivity of investigators even if they are independently funded, unless they have already achieved great renown at the time of the star's death.[29]

We also investigate whether the magnitude of the treatment effect varies with the collaborator's age at the time of death for the superstar. To do so, we interact the treatment effect with 8 indicator variables corresponding to different career age brackets: 5 to 10 years, 10 to 15 years, 15 to 20 years, 20 to 25 years, 25 to 30 years, 30 to 35 years, 35 to 40 years, and more than 40 years of career age at death. We then plot the corresponding coefficient estimates in Figure W3, along with their 95% confidence interval. The effect decreases monotonically with the age of the collaborator at death, becoming insignificantly different from 0 after twenty five years of career age. Therefore, researchers appear particularly vulnerable to the loss of a superstar coauthor in the early part of their scientific career.

Appendix X: Robustness Checks

In Table W8, we present the results of a number of robustness checks, using Model (1b) of Table III as a benchmark specification. In column 2, we examine whether a small number of stars with many collaborators drive the main results. We drop all collaborators for the 7 superstars with the highest number of collaborators (120 or more) from the estimation sample. The magnitude of the treatment effect drops only slightly, and remains highly statistically significant. In column (3), we add to the specification 10 indicator variables for the superstar's imputed career age. This decreases the magnitude of the treatment effect by from -0.086 to -.066. In columns (4a) and (4b), we explore the sensitivity of our results to changes in our arbitrary age cutoff for the the superstar's age at death. In (4a), we limit the sample to 71 stars who were 60 years old or younger at the time of their death. This results in an even higher magnitude for the extinction effect (-.113 instead of -.092). In contrast, we obtain a much smaller magnitude (-0.051) when we focus on the collaborators of 38 eminent scientists who die beyond the creative stages of their career — at 75 years of age or older (column 4b). This effect is also imprecisely estimated.

We then examine the possibility that our control group is contaminated because some of the control collaborators are separated from treated collaborators by a only few degrees in the coauthorship network. Specifically, we keep in the estimation sample only those scientists that are at least 3 degrees apart in the coauthorship network formed by all 10,349 superstars. These scientists represent 75% of the overall sample. In column 5, we find that the treatment effect increases in magnitude, which is consistent with the hypothesis that the effect of superstar extinction extends beyond the set of direct coauthors, but decays quickly with social distance.

Finally, we perform a small simulation study to validate the quasi-experiment exploited in the paper (column 6). We generate placebo dates of death for the control collaborators, where those dates are drawn at random from the empirical distribution of death events across years for the 112 extinct superstars. We then replicate the specification in Table III, column 1a, but we limit the estimation sample to the set of 5,064 control collaborators. Reassuringly, the effect of superstar extinction in this manufactured data (based on 500 replications) is a precisely estimated 0.

[29] As seen in Table W5, taken as a whole, the set of elite coauthors suffers a decline in output similar to the one observed for the universe of all coauthors (i.e., in Table III). At the risk of repeating ourselves, the elite sample is very heterogeneous, and does include young, old, and fading stars.

Appendix XII: Main Results for Anticipated Death

In Table W9 and Figure W4, we present some results pertaining to the 6,515 collaborators of 136 superstars who died prematurely, but whose particular circumstances imply that their passing was anticipated. The vast majority of these anticipated deaths are due to cancer. Since coauthors might alter their collaboration strategies even before the superstar's death, the case for exogeneity of the extinction event is weaker in this case.[30]

The results in Table W9 parallel exactly those presented in Table III. We find that the treatment effect is of lower magnitude than in the sudden case (especially when the estimation sample includes control collaborators), and less precisely estimated. We also find very little evidence of impact on the publication output without the superstar (columns 2a and 2b). Figure W4 mirrors Figure IIIA. The trajectory of output appears to begin its monotonic decline prior the superstar's death (though the corresponding interaction terms are very small in magnitude, and statistically insignificant). The treatment effect, though consistently negative in sign, reaches statistical significance only in the long run — 10 years or so after the superstar's death.

These findings suggest that collaborators and quite possibly the superstar him/herself adjust their behavior in anticipation of the star's impending death. Though the determinants and particular form of these endogenous responses are certainly worthy of study, they are beyond the scope of the present paper.

Appendix XII: An Alternative Interpretation Based on a Sociological Mechanism: Ascription

Sociological studies of the scientific reward system have provided some evidence supporting the existence of the "Matthew Effect,"[31] whereby scientists receive differential recognition for a particular scientific contribution depending on their location in the status hierarchy (Merton 1968; Cole 1970). It is possible that editors and reviewers ascribe positive qualities to research they are charged with evaluating because of the mere presence of the superstar's name on the authorship roster, regardless of the contribution's intrinsic merits.

The relevance of this dynamic in our setting is doubtful for two reasons. First, we observe a decline in the output written independently of the star (Table III); second, the treatment effect is not driven by the collaborators who have recent, or many collaborations; third, its onset is delayed until after the death of the star. These facts argue against an interpretation of the effect based on ascription.

[30]Most of the anticipated deaths are due to conditions with relatively short life expectancies; those with longer ones are not necessarily viewed as terminal until the final stages. Six scientists who died from a neurodegenerative disease constitute an exception. They were included in the sample because their obituaries implied they had remained actively engaged in research until a short period before their death. We verified that our results are robust to the omission of these six superstars.

[31] *"For unto every one that hath shall be given, and he shall have abundance; but from him that hath not shall be taken away even that which he hath"* [Matthew 25:29]

Chapter 3

For Art or for Money? Artistic Films and Star Careers in Hollywood

3.1 Introduction

Following the large literature on superstar economics [Rosen, 1981], much attention in labor economics, corporate finance, and related fields has recently focused on the compensation and career trajectories of star workers (Murphy [1999], Frydman [2005], Tervio [2003]). However, little is known about the role of non-pecuniary rewards in motivating star employees (notable exceptions are Stern [2004], Yermack [2004], and Rajan and Wulf [2006]). This gap is particularly glaring since non-pecuniary job characteristics may be particularly important drivers of performance for creative and knowledge workers, including most of the subjects of the superstar literature. This paper examines the role of artistic films in star careers in the American film industry, taking advantage of the wealth of public information on star careers and project characteristics in this setting. In particular, I use the well-known tension between artistic and commercial interests in the film industry to shed light on how the creative preferences of stars might drive their career trajectories.

Using data from all films released in the United States from 1980 and 2005 and the career histories of 100 star directors and 94 star actors, I document evidence on the interaction between artistic films and the value of stars over their careers. While there has been considerable debate over the value of stars in the film industry (Albert [1998], De Vany and Walls [1999], Elseberse [2006], Ravid [1999]), simple estimates from my results suggest that stars are associated with additional revenues of \$70-\$110 million, exceeding their salary costs of \$20-30 million. Artistic films make up 12% of star careers, and they are associated both with significantly lower film revenues and lower monetary compensation. The propensity for stars to work on artistic films is relatively constant across their career, although it is slightly higher when stars are under 30 or over 60 relative to middle age. Furthermore, artistic films are significantly associated with Oscar awards.

The willingness of stars to take substantial pay cuts to work on artistic films suggests that these films confer non-pecuniary benefits. However, artistic films could also be related to human capital formation or the signaling of star ability. Artistic films may help stars acquire skills that increase their value for later films. Alternatively, artistic films may help stars reveal their ability to studios and thus improve future career prospects. However, both human capital formation and ability signaling would suggest that artistic films are mainly valuable early in careers. In contrast, I find that stars continue to spend a significant fraction of their time on artistic films after they have achieved wide recognition and even toward the end of their careers. While I find no evidence that artistic film experience contributes to future salaries or film revenues, artistic films do increase the chances of current and future Oscar nominations, a plausible proxy for the non-pecuniary benefits that are important to stars.

This paper provides suggestive evidence that a preference for artistic films may shape the compensation and careers of stars in the film industry. However, because only completed film choices and a limited sample of compensation data can be observed, the empirical setup cannot definitively eliminate alternative explanations. Stars may work on artistic films during times when they have few commercial alternatives. Artistic recognition could also lead to unobserved forms of remuneration such as advertising contracts. Even if stars indeed have a preference for artistic films, the incentives for studios to produce these films are unclear. Ultimately, the results presented in this paper provide a starting point for addressing the economic role of artistic films in the film industry.

3.2 Sample Description

My primary source of data is Box Office Mojo, the leading internet resource for data on film revenues. The database includes a comprehensive set of 6632 films released from January 1980 to December 2005 in the United States, with data on domestic and international revenues[1]. The comprehensive set of films released from 1980 to 2005 is supplemented by the complete career histories of films made by star actors and directors prior to 1980 (see Section 3.2.2). I obtain data on each film's production and distribution companies, MPAA rating, and genre from the Internet Movie Database (IMDB).

3.2.1 Definition of artistic films

The tension between the commercial interests of firms and the creative interests of talent in the film industry has been widely discussed both in academic studies and the popular media (Fee [2002], Biskind [2004], Waxman [2005], Mottram [2006], Goldman [2000], McCabe [August 20, 2004]), motivating my focus on artistic films as potential sources of nonpecuniary rewards for stars. Based on industry accounts (Biskind [2004], Waxman [2005], Mottram [2006]), I use two definitions of artistic films: 1) films produced by artistic divisions of large studios and 2) films exhibited at the Sundance Film Festival.

The first definition uses the organizational structure of the major film studios. Each of the seven major studios [2] consists of numerous divisions that specialize in different types of films or that resulted from previous acquisitions. Based on Biskind [2004], Vachon and Bunn [2006] and other industry accounts, I designate 15 of the 35 subsidiaries as artistic. Although this classification is somewhat subjective, industry insiders generally regard artistic orientation as roughly binary in nature and generally concur on the set of artistically-oriented subsidiaries. In addition, many of these subsidiaries employ identifying adjectives such as "classics" and "independent" in their names to signal their artistic orientation.[3] The artistic subsidiaries both produce and distribute films, often distributing films that are independently-produced. Since subsidiaries distribute all films they produce and also purchase rights for independently-produced films, I classify artistic films by distributor in order to capture a larger sample of films. However, all results are similar if only subsidiary-produced films are considered.

[1] Revenues from regular releases are pooled with those from re-releases, director's cuts, and other special showings.

[2] See Appendix A for a list of the studios and subsidiaries

[3] As an indication of the artistic subsidiaries' reputation for artistic freedom, a former executive of United Artists Classics stated that "Once the script and the director were set and it was clear the movie could be made for the budget they wanted, then we stepped aside and let the artists do their work Biskind [2004, p. 18])."

The second definition of artistic films includes all films that have been shown at the Sundance Film Festival, the largest independent film festival in the United States that has been running since 1985 [4]. Admission to the festival is competitive and judged on artistic merit, and the only Sundance films included in my sample are the small subset that are purchased by distributors and later released theatrically.

The first artistic film definition includes only films distributed by one of the major studios, including most of the highest-profile artistic films. In contrast, the second is based on a universe of films so esoteric that most of them are never even released. Together, the two definitions encompass a wide spectrum of artistic film types and allow comparison between the two ends of the spectrum.

3.2.2 Definition of stars

While labor market in the film industry includes a seemingly endless supply of aspiring talent, only a few top stars are perceived to be able to reliably draw audiences and generate value for studios. Because stars are able to command high pecuniary wages and their ability is already widely recognized, the non-pecuniary value of artistic films can be more-readily distinguished from their use as a signal of ability.

Previous studies have used past Oscar nominations, critical reviews, and film revenues to identify stars (Albert [1998], Elseberse [2006], Ravid [1999], Ravid and Basuroy [2004]). However, since this study focuses on the artistic preferences of stars, using measures that are directly correlated with artistic ability such as awards and critical reviews would be problematic. Moreover, the economic performance-based measures such as past revenues do not attempt to disentangle a star's individual contribution from the many other inputs that contribute to a film's success.

Following [De Vany and Walls, 2002b] and [De Vany and Walls, 1999], I use a broad definition of stardom based on the *premier 100*, an annual list compiled by premier magazine of the most powerful players in Hollywood. The *Premier 100* provides a comprehensive index that represents an aggregation of information about a star's past performance by industry without explicitly relying on either artistic or commercial metrics. Because many of the most powerful players in Hollywood are not stars but agents, executives, and producers, stars represent only a fraction of the Premier 100, leaving a small universe of star actors and directors.

I identify 100 directors and 94 actors who have appeared on the Premier 100 in any year. As of 2005, the professional organizations Director's Guild of America and the Screen Actor's Guild of America listed memberships of about 12,000 and 120,000 members, respectively [5], so stars represent a tiny fraction of the labor pool.

3.2.3 Descriptive Statistics

Panel A of Table 3.1 shows descriptive statistics for 6632 films released in the U.S. between January 1, 1980 and December 31, 2005. As shown in the last column, film revenues exhibit an extremely skewed distribution, with mean domestic revenues of $26 million and worldwide revenues of $52 million and median revenues of $ 6 and $ 7 million. The skewed and risky nature of returns in the film industry has been documented previously by De Vany [2004], Ravid [1999], and others. Moreover, comparison between domestic and worldwide revenues indicate that only the most popular films have significant overseas revenues. Production budgets are available for only 2636 films, and are more likely to be available for higher-budget films. While the mean (median) production budget is $32 (21) million, the

[4] The festival was known as the Utah/US Film Festival starting in 1978, and in 1985 Robert Redford's Sundance Institute took over management of the festival. The festival's name was changed to the Sundance Film Festival in 1991. The film festival and its sponsor, the Sundance Institute, aims to support films which maintain the artistic integrity of film-makers without being influenced by studio managers (See www2.sundance.org)

[5] See http://www.dga.org and http://www.sag.org.

worldwide revenues for films with a known budget is \$81 (25) million. Seven percent of films are nominated for an Oscar in one of the eight major categories. Ten percent have a star director, and 24% have a star actor.

In the first three columns, I show film characteristics for films distributed by a major studio and films in the two artistic film categories. The first column shows films distributed by one of the seven major studios but *not* included in either of the two artistic categories, which I will call "studio" films for simplicity. Although they comprise only 54% of all films released, studio films account for 89% of worldwide revenues. In addition, studio films are more likely to receive Oscar nominations (8%) and employ star actors (16%) and directors (35%) than the average film in the sample.

The second and third columns show firm characteristics for subsidiary and Sundance films, with the second column excluding films that fall under both categories, and compare these characteristics with those of studio films. Subsidiary films have mean (median) domestic revenues of \$ 10 (\$ 2) million and worldwide revenues of \$18 (\$ 3) million, while Sundance films have substantially lower mean (median) domestic revenues of \$ 6 (\$ 1) million and worldwide revenues of \$10 (\$ 1) million. Means and median revenues for both types of artistic films are significantly different from those of studio films at the 1% level. While subsidiary films are 3% more likely to obtain Oscar nominations compared with studio films, Sundance films are 2% less likely, with the differences significant at the 1% and 5% levels. Both types of artistic films are also only about half as likely to include stars as studio films.

Panel B provides summary statistics on the careers of the 100 star directors and 94 star actors observed in the sample. Most careers are truncated in the sample. As of 2005, both the mean and median director age are 54, ranging from 34 to 77. Actor ages in 2005 have a mean of 45, median of 43, and range from 19 to 75. The mean rank on the *Premier 100* (calculated across all years when stars appear on the list) is 86 for directors and 84 for actors. Although a minority of studio films contain stars as shown in Panel A, both actors and directors spend about 90% of their careers on films distributed by one of the major studios. Directors spend 13% of their careers on subsidiary films and 7% on Sundance films, and they receive Academy Award nominations for Best Director for 8% of their films. Similarly, actors spend 10% of their careers on subsidiary films and 5% on Sundance films, and receive acting nominations for 4% of their films.

3.3 Overview of star careers

Figure 3-1 shows the distributions of star ages in 2005. As shown in Table 3.1b, most stars are still career active at the end of the sample. While the sample includes some stars at the beginning and end of their careers, the bulk are middle-aged, suggesting a rise and then fall in star power across a career. While the distribution of directors tends to be older than actors, the similarity in the overall shapes of the distributions suggest that the variation in star power over the career lifecycle is similar for the two star types.

In contrast, the distribution of total career films differs substantially across directors and actors as shown in Figure 3-2. While total career films are highly skewed for directors, the hump-shaped distribution for actors roughly mirrors the distribution of age in 2005. Thus, comparing Figures 3-1a and 3-2a suggests that director productivity is more skewed than actor productivity. One potential explanation is that lower-ability directors are not rehired for subsequent films, as found by John et al. [2002] for a sample containing both star and non-star directors starting their careers in the mid-1980s. Alternatively, there could be differences in productivity independent of director ability.

Figure 3-3 shows how productivity varies along the lifecycle. Directors make about 0.4 films per year, with productivity declining across their careers. In contrast, actor productivity exhibits a hump-shaped pattern that increases from ages 15 to 35 and begins to decline after age 40. Productivity is observed to increase past age 60 for both directors and actors, but this effect may be due to a compositional difference between old and young stars in the sample.

Figure 3-4 shows cross-sectional differences in the fraction of artistic films in star careers. The vast majority of stars do no artistic films in their careers, while those that do ar least one artistic film vary substantially in the percentage of artistic films in their careers. Since actors work on many more films during their careers, the fraction of artistic films in acting careers is unsurprisingly less skewed than for directors.

To show how film types and remuneration evolve along the career path, Table 3.2 shows the percentage of artistic films and Oscar nominations and mean salaries and revenues by age group. Although the propensity to work on subsidiary films fluctuates across the career, the percentage of subsidiary films exceeds 10% for directors between ages 30 to 34, 60 to 64, and over 70. Similarly, actors are most likely to participate in artistic films between the ages of 20 and 29 and above 60. In contrast, both actors and directors are most likely to work on Sundance films under the age of 30. However, while directors over 55 never work on Sundance films, the percentage of Sundance films levels out for actors at about 3% after age 30. The U-shaped pattern of artistic film choice across the career is even more striking for both directors and actors conditional on doing at least one artistic film in their careers (not shown in the table).

Oscar chances generally increase with age for both types of stars, although star directors who do their first films before the age of 25 are also very likely to gain nominations. Star directors' Oscar chances range from 2% between ages 25 to 29 to 14% between ages 60 to 64 and over 70. Oscar chances for star actors range from 0% for stars under 15 years older to 9% between ages 60 to 64. Compensation for both directors and actors start at less than $5 million per film under age 30. For directors, mean salary increases to $53 million between ages 45 and 49 before dropping to $10 million between ages 50 to 54. Between ages 55 and 59, the mean compensation of $137 million is driven by George Lucas's $400 million take for *Star Wars I: The Phantom Menace*[6].

The revenues of their films can be used as an initial measure of stars' economic value, although this measure is imperfect since it does attempt to disentangle a star's contribution from other factors that impact revenues. In Section 3.6, I will construct more detailed measures based on an individual star's marginal contribution that control for other firm characteristics. As shown in the second-to-last column of the table, while the initial films of directors tend to be modest with mean revenues of $33 million for directors under 25, average revenues are between $100 and $200 million until age 64, peaking at an average of $201 million between ages 45 and 49 and declining to $79 million after age 70. Child actors are associated with very high revenues, averaging $104 million between the ages of 5 to 9 and $141 million between 10 and 14. As actors enter adolescence, their film revenues decline to $61 million from age 15 to 19 and then climb monotonically to a peak of $ 140 million between ages 40 to 44. After age 45, revenues for actors fluctuates between $97 and $140 million.

As a measure of the division of rents between stars and studios, the salary to revenues ratio for directors increases nearly monotonically from less than 1% between ages 25 to 29 to 27% between ages 60 to 64 before dropping to 18% between ages 65 to 69. The hump-shaped pattern of rent-sharing is similar for actors, whose salary to revenues ratio increases from less than 10% under age 30 to 17% between ages 45 to 54, dropping to 15% between 55 and 64 and 12% over age 65 [7].

A few key patterns emerge from the evidence in this section. The compensation, film revenues, and bargaining power of stars displays a hump-shaped pattern that peaks at middle age, suggesting that human capital accumulates and then depreciates across the lifecycle[8]. The probability of obtaining Oscar nominations generally increases with

[6]Since only 41 compensation observations are available for star directors and are more likely to be available for better-known films and directors, conclusions about compensation should be tempered by concerns over potential selection bias.

[7]These figures exclude two extreme outliers: Sylvester Stallone's 2002 film *D-TOX* for which he made $20 million but which garnered only $83,000 in revenues, and James Van Der Beek's 2001 film *Texas Rangers* which earned him $3 million and made $670,000 in revenues

[8]Human capital in this setting could take the form either skills related to directing and acting or popularity with

age, suggesting that artistic ability does not suffer as much depreciation as commercial appeal. In contrast to proxies for stars' economic power, the propensity to work on artistic films is highest during the early and late stages of the career. Both the human capital accumulation and ability signaling explanations for why stars work on artistic films would predict that stars should do artistic films mainly during the early part of their careers. Thus, the evidence for continuing participation in artistic films during middle age and beyond suggests that stars are also responding to an intrinsic taste for artistic films. Nonetheless, only a small percentage of stars work on any artistic films in their careers, suggesting that the taste for these artistic films is heterogeneous.

3.4 Star compensation

This section examines the determinants of star compensation, focusing on the relationship between pecuniary rewards and artistic films. The IMDB lists salary information for a total of 495 films by 24 star directors and 74 star. Table 3.3 shows OLS regressions of log compensation on characteristics of the star, the type of role, and film. Nineteen stars have only one film in the compensation sample, but star fixed-effects are included for the 79 stars with more than one film in the sample. Unreported controls are also included for film rating (G, PG, PG-13, and R), genre (action, comedy, drama, romance, and thriller), and the job(s) of the star in the film (voice actor, actor, producer, executive producer, director, and writer).

Salaries have significantly increased over time. The coefficient on a year trend indicates that for each year between 1980 and 2005, star salaries increased by 11%. Coefficients on film count and age are both positive and significant with negative and significant quadratic terms, confirming the concave path of compensation across the lifecycle shown in Section 3.3. While a move from age 25 to 30 is associated with an 360% increase in compensation at the mean film counts for these ages, a similar move from 60 to 65 is associated with an 2% decrease in compensation.

Conditional on film characteristics, star fixed-effects, and career stage, stars receive substantially higher compensation for films made by the major studios and lower compensation for artistic films. Studio films are associated with 88% higher compensation, while subsidiary and Sundance films are associated with 69% and 58% lower compensation.

The results are consistent with stars giving up monetary compensation for the nonpecuniary rewards of working on artistic films. One caveat to these results is that compensation is available for only a fraction of star-film pairings. The IMDB lists compensation estimates based on media reports, so data are more likely to be available for higher-profile films and higher compensation levels. Since artistic films generally receive less media attention, selection bias is likely to understate the negative relationship between compensation and artistic films.

It is also possible that unobserved variation in work opportunities drive the lower compensation for artistic films. If stars only take on artistic films when they have few commercial alternatives, then their lower compensation may not reflect a compensating differential but rather lower opportunity costs. It is difficult to completely disprove this alternative since simultaneous job and wage offers are not available. However, in nearly all cases in which stars accept artistic films, studio films are produced with lower-powered actors and directors which would likely have resulted in higher monetary compensation for stars. Thus, it is unlikely that the reduced compensation associated with artistic films is due solely to the unavailability of commercial work.

audiences.

3.5 Oscars, artistic films, and star careers

An observable proxy for the nonpecuniary benefits that stars may receive from artistic films is the attention associated with the Academy Awards, the most prestigious honor in Hollywood. In Table 3.4, I examine the relationships between artistic films and Oscar nominations in the eight major categories (best picture, best director, best actress/actor, best supporting actress/actor, and best adapted/original screenplay). I omit Oscars outside of the major categories since these are often for specific components of films such as costumes, music, and special effects that stars are both unlikely to contribute to and unlikely to benefit from.

All of the specifications include dummies for film genre and rating. As a benchmark, Column (1) presents a logistic regression for whether a film receives any Oscar nomination using the comprehensive film sample from 1980 to 2005. Computed at the means of the other covariates, studio films are 2.6% more likely to be nominated nomination that is significant at the 1% level. Subsidiary films are associated with a 1.3% higher probability of nomination that is significant at the 5% level, and Sundance films are associated with an insignificant 0.5% lower probability of nomination. Star directors contribute much more to a film's Oscar chances than star actors. The presence of star directors and actors are associated with 7.5% and 1.5% higher nomination probabilities, respectively, and both coefficients are significant at the 1% level.

The sample for specification (2) includes only films by stars, and the model includes controls for star age and age squared, film count, subsidiary film count, and Sundance film count. The sample is supplemented by all films with a star actor and director released prior to 1980. Career statistics are averaged for films with multiple directors or actors. Since nearly all Oscar winners with stars are distributed by large studios, the *Studio* dummy is omitted from this specification. The career film count variables capture how experience with different types of films contribute to Oscar chances. The subsidiary effect is stronger in specification (2) compared with specification (1), accounting for an 8.5% higher probability of nomination. The coefficient for Sundance films is larger than in specification (1) with a negative 2.5% marginal effect, but it remains insignificant. Consistent with the lifecycle variation in Oscar recognition described in Section 3.3, the coefficients on director and actor age are both positive, with negative coefficients on their quadratic terms. The director age coefficients indicate that decade increase in age is associated with a 4.1% higher nomination probability computed at the means of the other variables. The actor age coefficients are much smaller and insignificant. The coefficients on the career film count variables are all insignificant in specification (2), although their magnitudes are economically significant. Their signs indicate that both overall film experience and Sundance film experience improve nomination chances, while director subsidiary experience decreases chances and actor subsidiary experience improves the chances.

Specifications (3) and (4) attempt to isolate the role of artistic films and star career experience for stars' chances of garnering Oscar nominations for themselves. The sample for specification (3) consists of all films with a star director. The coefficient on subsidiary films is significantly larger than in specifications (1) and (2). Subsidiary films are associated with a 15% higher chance of being nominated for Best Director. The magnitude of the coefficient for Sundance films is also larger, although it remains insignificant. The coefficients on director age and director age squared are smaller in magnitude and insignificant in this specification. However, director film count is marginally significant at the 10% level with a marginal effect of 0.25% for every film. The coefficients on director artistic film count are both small and insignificant.

Specification (4) focuses on nominations for Best Actress, Best Actor, Best Supporting Actress, and Best Supporting Actor, and the sample is restricted to films that include star actors. The sample also includes multiple observations for the same film if more than one actor is nominated. Subsidiary films no longer significantly contribute to nomination chances for the acting prizes, although the coefficient on Sundance films becomes marginally significant at the 10% level with a -1.1% marginal effect. The coefficients on actor age and career film counts are also larger in magnitude with the exception of film count and subsidiary film count. The coefficient on actor age is significant at the 10% level and together with the quadratic age term, indicates that a decade increase in actor age is associated with a 2.6% higher probability of nomination. The coefficient on Sundance film count is more than twice as large compared with specification (2) and is significant at the 1% level, indicating that one additional past Sundance film is associated with a 0.35% higher nomination probability.

Overall, the results indicate that subsidiary films are significantly more likely to receive Oscar nominations, particularly Best Director nominations for star directors. Thus, Oscars are a tangible source of the non-pecuniary rewards that stars obtain from working on subsidiary films that may compensate for their lower monetary compensation. Since Sundance films are negatively associated with Oscar nomination possibly due to their small audience size and limited visibility to Academy voters, the prestige of the Oscars cannot account for why stars work in Sundance films. While Oscars are ostensibly awarded for individual films, industry insiders commonly note that awards are also partially based on career achievement and visibility. Age and past film experience are positively correlated with nomination probabilities, providing evidence either that artistic ability increases with experience or that voters are more likely to recognize ability for established stars. Table 3.4 provides only limited evidence that past artistic film experience contributes to nominations, although Sundance film experience is positively and significantly associated with nominations for star actors.

3.6 Determinants of film value

In this section, I use a framework similar those used in previous studies (Litman and Kohl [1989], De Vany and Walls [1999], Ravid [1999], and John et al. [2002]) to examine how artistic films relate to the economic value of films and stars. I use worldwide film revenues as the main proxy for film value, and Table 3.5 presents results from OLS regressions of log worldwide revenues on film, studio, and star characteristics. Box office revenues are the primary focus of film executives (Ravid and Basuroy [2004]), although significant components of both film costs (e.g. production and advertising budgets, overhead, and film duplication expenses) and revenues (e.g. television, DVD sales, merchandizing) are incomplete or unavailable. These omissions are likely to understate the value of the most successful films, since blockbusters are the films most likely to profit from ancillary revenue sources. Although the dataset includes some data on production and marketing costs, they are available for only 41% and 7% of the sample, so I omit these variables in order to more precisely identify the coefficients. However, the results are consistent when production budget is included. Results are also qualitatively similar when considering film returns (revenues divided by costs) as in Ravid [1999] and De Vany and Walls [2002a].

All specifications in Table 3.5 include fixed-effects for film genre and rating. The first specification uses the comprehensive sample of 6499 films from 1980 to 2005 to provide a benchmark estimate for film type and to quantify the relationship between stars and film revenues. The estimates indicate that studio films are associated with 730% higher revenues compared with non-studio films. Subsidiary and Sundance films are associated with 52% and 31% lower revenues. Dummy variables for whether the film includes a star director or actor indicate that stars are associated with 270% and 250% higher revenues. The coefficients for studio, subsidiary, and Sundance films and for star director and star actor are all significant at the 1% level. While previous studies have emphasized the unpredictability of film profitability (Ravid [1999], De Vany and Walls [1999]), the R^2 of 0.52 indicates that a large portion of the variation in film revenues can be explained by a publicly-observable set of film characteristics.

The remaining specifications focus on films with stars, including all films by stars that were released prior to 1980. Specification (2) includes star fixed-effects, quadratics in director and actor age, and career film counts as described in Section 3.4. The coefficient for studio films is smaller than in specification (1), while the coefficients on subsidiary and Sundance films are larger in magnitude. Among films with stars, studio distribution is associated with 440% higher revenues, while subsidiary and Sundance films are associated with 61% and 55% lower revenues. Similar to the results for compensation and Oscar nominations (Sections 3.4 and 3.5), both director and actor age positively related to revenues with negative quadratic terms, although only the coefficient on actor age is marginally significant at the 10% level. The point estimates suggest that a move from age 20 to 25 is associated with a 3.3% revenue increase for star directors, while a move from age 60 to 65 is associated with a 15% revenue decrease (conditional on the mean film counts at each age). For actors, a similar move from age 20 to 25 is associated with a 9.6% revenue increase while a move from age 60 to 65 is associated with a 1% revenue increase. The coefficients on artistic film counts are insignificant but large in magnitude. Each past subsidiary film is associated with a 12% revenue increase for directors and 7% revenue increase for actors. Each past Sundance film is associated with an 18% revenue decrease for directors and 8% revenue increase for actors. The 0.84 R^2 for specification (2) indicates that including star fixed-effects results in a highly predictable model for film profitability.

Specifications (3) through (5) replicate specification (2), splitting the sample into mutually-exclusive categories by commercial and artistic orientation. The sample in specification (3) includes only films distributed by the major studios that do not fall into either artistic film category, specification (4) includes subsidiary films that are not shown at Sundance, and specification (5) includes all Sundance films. Only specification (3) includes star fixed-effects since few stars participate in more than one subsidiary or Sundance film.

The signs of the coefficients for director and actor age and film count are very consistent across specifications (3) through (5), although their magnitudes can vary widely. Director and actor age have the largest effects on artistic films, particularly Sundance films. Unconditional on film counts, a move from age 20 to 25 for directors is associated with an 9% increase in revenues for studio films, a 7% increase in revenues for subsidiary films, and a 31% increase in revenues for Sundance films. Similarly, a move from age 20 to 25 for actors is associated with 6%, 19%, and 23% increases in revenues for studio, subsidiary, and Sundance films. Moves from age 60 to 65 are associated with 0% 41%, and 57% declines in revenues for directors, and 16%, 2%, and 31% declines for actors. Director age is significant at the 5% level for subsidiary and Sundance films, while director age squared is significant at the 10% level for subsidiary films. Actor age and age squared are only significant for studio films, at the 1% and 5% levels. The star film count variables are generally insignificant with the exception of actor subsidiary film count in specification (3), which is significant at the 5% level and indicates that one additional past subsidiary film is associated with a 14% increase in revenues. The signs of the star film count variables generally indicate that director film experience is positively correlated with revenues only for artistic films, while actor film experience is positively correlated with revenues for all film types. Subsidiary film experience is generally positively correlated with revenues for all film types, while Sundance film experience is inconsistently related to revenues.

Table 3.5 reveals several key pattern. Artistic films garner substantially lower revenues, particularly for films with stars. This result helps to explain why studios offer lower compensation to stars for working on artistic films as shown in Section 3.4, while a preference for artistic films can help explain why stars are willing to accept lower compensation. Although subsidiary film experience by stars is weakly related to higher future revenues, the results provide limited evidence that artistic films are a useful form of human capital investment with respect to film value.

The results also contribute to the literature on the value of stars, with simple calculations indicating that stars confer substantially more revenues than they earn as compensation. On average, films with star directors earn worldwide revenues of $158 million. The coefficient from specification (1) indicates that a similar film with the same characteristics would have only earned $42 million without the star, yielding a marginal value of $116 million. Similarly, the marginal value of star actors can be estimated at $73 million. Mean salaries are $27 million for directors and $10 million for actors, suggesting that stars provide net value to studios. This finding is consistent with the conclusions of

Albert [1998], Basuroy et al. [2003], Litman and Kohl [1989], and Elseberse [2006], while running counter to those of De Vany and Walls [1999] and Ravid [1999].

Although the results provide support for the notion that stars add value to films, they may overestimate the causal effect of stars. If stars provide a multiplicative effect on film revenues, studios are likely to systematically select them for the films most likely to succeed. The star effects may thus be biased upward due to correlation with unobserved heterogeneity in the expected success of films. However, selection is unlikely to explain the entire effect. In particular, star directors are often the writers and key creators of the intellectual property on which films are based, making their participation critical to the film's very existence. Nonetheless, variation in film revenues across the lifecycle and by film experience may also be partially driven by unobserved differences in film quality as opposed to changes in star value or ability.

3.7 Artistic films and star heterogeneity

In this section, I explore heterogeneity in star types by comparing correlations between measures of stars' artistic orientation and their economic value. Pairwise correlations between career characteristics measured across the entire career. When possible, the measures attempt to disentangle artistic orientation and star value from star age, which is important concern since most careers are significantly truncated in the sample.

The first three measures in Table 3.6 are proxies for artistic orientation: the percentage of subsidiary films, Sundance films, and films in which the star is nominated for an Oscar across their careers. For directors, % subsidiary films is positively and significantly correlated with both % Sundance and % Oscar nominated at the 1% and 5% levels, with correlations of 0.29 and 0.25. However, % Sundance and % Oscar nominated have an insignificant correlation of -0.07. For actors, all three measures are positive correlated with each other, although only the correlation between % subsidiary and % Sundance is significant at 0.32 at the 1% level.

The results are similar for measures of the total number of artistic and Oscar-nominated films or alternatively, 0/1 variables indicating at least one artistic or Oscar-nominated film in the career. However, these measures are more highly correlated with star age in 2005, so I present measures based on percentages.

The remaining five measures capture different dimensions of star value. *Mean compensation* is a star's average compensation across her career. *Mean rank* is 101 - the star's mean rank in the *Premier 100* across all of the years in which the star appears on the list. The mean rank measure is inverted for ease of comparison so that star power increases with rank. *Adjusted film count* captures star productivity conditional on age, and it computed as a star's total film count in 2005 subtracted by the mean film count within 5-year age bins for star age in 2005 (the adjustments are performed separately for directors and actors).

I use two measures of star value based on film revenues. *Mean revenues* is the raw average of worldwide revenues for all films in a star's career. *Revenue fixed-effect* is the star's fixed-effect from Specification (2) in Table 3.5, which disentangles the star's marginal contribution to film value from other film characteristics. As shown above, artistic films are significantly negatively correlated with both star compensation and film revenues, so artistic orientation could mechanically lead to negative correlations with the measures of star value. However, the results are very similar when basing mean salary, mean revenues, and revenue fixed-effects only on *non-artistic* films.

The five measures of star value are generally positively correlated with each other, with positive pairwise correlations in 7/10 cases for directors and 8/10 cases for actors. For directors, the only significant correlations those between mean salary and mean revenues (0.63) and revenue fixed-effects and mean revenues (0.44), which are both significant at the 1% level. For actors, mean salary is significantly correlated with mean rank (0.50, 1%), adjusted film count (0.21, 10%), mean revenues (0.49, 1%), and revenue fixed-effects (0.26, 5%). Mean revenues are also positively correlated with mean rank (0.51, 1%) and revenued fixed-effects (0.30, 1%).

Correlations between artistic orientation and star value are generally negative, with 10/15 negative correlations for directors and 9/15 for actors. Three correlations are significant for directors: mean revenues with % subsidiary (-0.2, 5%), mean rank with % Sundance (-0.22, 5%), and revenue fixed-effects with % Oscar nominated (0.31, 1%). Four correlations are significant for directors: mean salary with % subsidiary (-0.29, 5%), mean salary with % Sundance (-0.30, 1%), mean rank with % Sundance (-0.25, 5%), and revenue fixed-effects with % Subsidiary (0.28, 1%).

Overall, the results support the conclusion that stars who work on artistic films sacrifice both monetary compensation, film revenues, and bargaining power. Artistic orientation is also weakly correlated with lower age-adjusted productivity, suggesting that artistically-oriented stars either choose to be less prolific or receive fewer film offers. However, these effects may be mitigated if they garner Oscar attention for high artistic achievement. While the proxies used in this section are crude, and measurement error due to career truncation is impossible to fully eliminate, the results provide evidence that heterogeneity in star orientations toward artistic films is a significant driver of career trajectories.[9]

3.8 Conclusion

This paper presents evidence that star directors and actors sacrifice significant amounts of monetary compensation to participate in artistic films. While artistic films could serve to provide stars with skills that are valuable for future films or to signal star ability, these explanations cannot account for the continued participation in artistic films for mature stars and the U-shaped pattern in the propensity to work in artistic films across star careers that I document. Furthermore, I find little evidence that artistic films are associated with higher future compensation for stars.

Honor and prestige from Oscar nominations are one dimension of the non-pecuniary rewards stars receive from artistic films. Films made by artistic subsidiaries of major studios are significantly positively correlated with Oscar nominations, but Sundance films are negatively correlated. Thus, prestige may be only one component of these non-pecuniary rewards. The taste for artistic films also appears to be heterogeneous among the population of stars.

This analysis leaves open the question of why large studios choose to invest in artistic films. As suggested by Fee [2002], agency problems and conflicts of creative control suggest that artistic films may be better suited for independent production and distribution. However, studios may find it beneficial to use artistic films as a way to attract talent in a free-agent market with a small set of marketable talent. Alternatively, the substantial investments in artistic films made by major studios could be evidence for agency problems between studio owners and artistically-oriented executives.

[9]Another interesting finding from Table 3.6 is that both star ranks and salaries are more highly correlated with mean revenues than revenue fixed-effects, while % Oscar nominations are more highly correlated with revenue fixed-effects than mean revenues. Thus, industry insiders do not seem to sufficiently filter stars' distinct contributions to film value from raw revenues.

Table 3.1
Summary Statistics
Panel A: Film characteristics

Characteristics for 6632 films released in the United States between 1980 and 2005. Star directors and actors are defined as those who appear on the Premier 100 list of the most powerful people in Hollywood in any year between 1990 and 2005. A film is considered Oscar-nominated if it receives a nomination in one of the eight major categories: best picture, best director, best actress/actor, best supporting actress/actor, and best adapted/original screenplay. *Studio* films are those distributed by one of the seven major studios (see Appendix A). *Subsidiary* films are those distributed by an artistic subsidiary of one of the seven major studios. *Sundance* films are those that have been shown at the Sundance film festival. In the table, *Subsidiary* and *Sundance* films are excluded from the *Studio* category, and *Sundance* films are excluded form the *Subsidiary* category, making all of the categories mutually exclusive. ***,**,* denote that the mean (median) difference between the column category and the *Studio* category is significant at the 0.01, 0.05, and 0.10 levels based on a t-test (Wilcoxon rank sum test).

	Studio	Subsidiary	Sundance	All
N	3562	825	654	6632
% of films	54%	12%	10%	100%
Domestic revenues ($M)	42.12	9.96 ***	6.20 ***	25.62
	(21.49)	(2.15) ***	(0.85) ***	(5.92)
Worldwide revenues ($M)	68.82	18.39 ***	9.69 ***	41.66
	(23.29)	(2.84) ***	(0.87) ***	(6.91)
Production budget ($M)	42.91	14.64 ***	7.30 ***	32.22
	(33.20)	(9.31) ***	(3.21) ***	(21.32)
Oscar nominated	0.08	0.11 ***	0.06 **	0.07
Star director	0.16	0.08 ***	0.04 ***	0.10
Star actor	0.35	0.19 ***	0.11 ***	0.24

Panel B: Star careers

Career statistics for star directors and actors. *Mean rank* is a star's mean rank in the *Premier 100* across all of the years in which she appears on the list, where 1 is the most powerful position and 100 is the least powerful. *Fraction studio*, *Fraction subsidiary*, and *Fraction Sundance* indicate the percentage of a star's total films that are distributed by a major studio, distributed by an artistic subsidiary of a major studio, or appeared at the Sundance film festival. *Fraction Oscar nominated* indicate the fraction of films for which star directors receive a Best Director nominations or star actors receive Best Actress, Best Actor, Best Supporting Actress, or Best Supporting Actor nominations.

	Directors N = 100				Actors N = 93			
	Mean	Median	Min	Max	Mean	Median	Min	Max
Age in 2005	54	54	34	77	45	43	19	75
Career film count	8	6	1	34	25	25	1	63
Mean rank	86	93	10	100	84	87	46	100
Fraction studio	0.91	1	0	1	0.90	0.90	0	1
Fraction subsidiary	0.13	0	0	1	0.10	0.07	0	0.39
Fraction Sundance	0.07	0	0	1	0.05	0.03	0	0.33
Fraction Oscar nominated	0.08	0	0	1	0.04	0	0	0.34

Figure 3-1
Distribution of Star Age in 2005

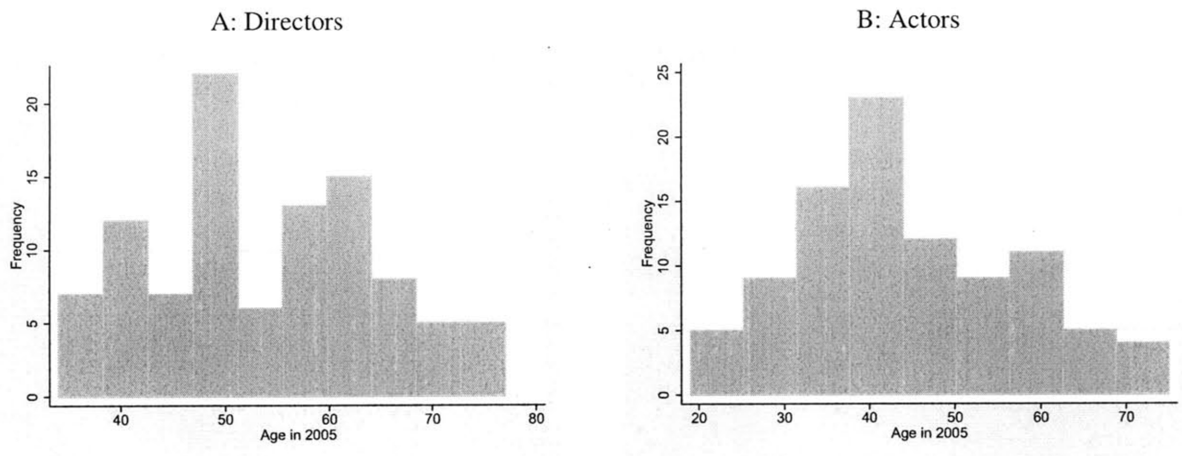


Figure 3-2
Distribution of Career Film Count

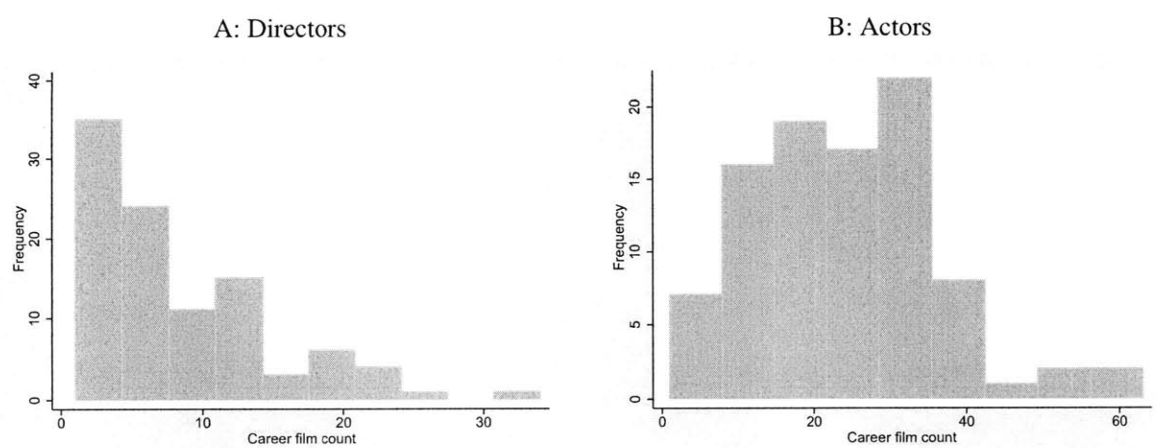


Figure 3-3
Star Productivity Over the Life Cycle

Circles show the mean number of films released each year for stars by age. Bars represent standard errors above and below the mean. Data points are omitted if fewer than five stars are observed at that age.

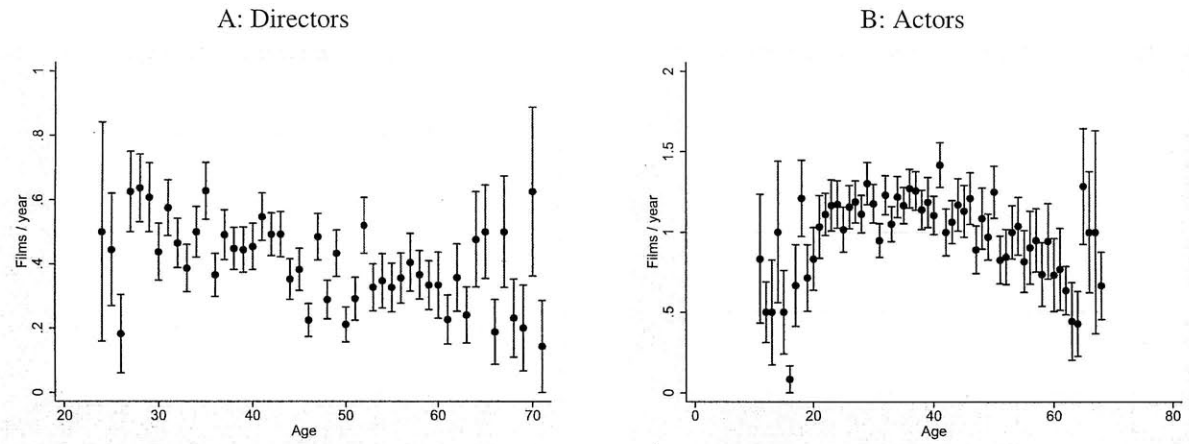


Figure 3-4
Distribution of Artistic Film Preference

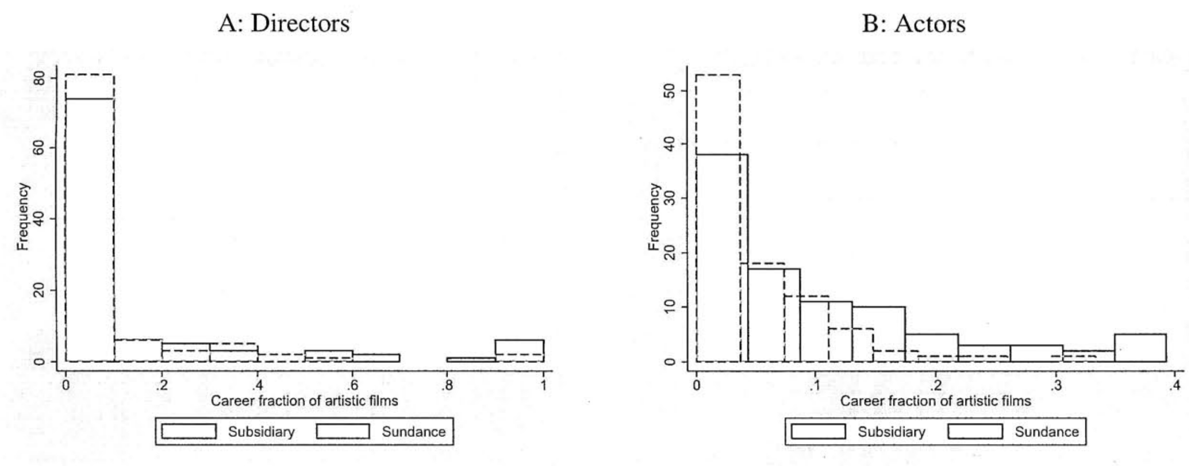


Table 3.2
Film Characteristics and Compensation Across the Career

Each row shows statistics for films released when stars are within the given age range. *% Star Oscar nominated* indicates films for which star directors receive Best Director nominations or star actors receive Best Actress, Best Actor, Best Supporting Actress, or Best Supporting Actor nominations. *Mean compensation* and *Mean worldwide revenues* are computed per film, and *Mean compensation / revenues* is computed across films with data on both star compensation and worldwide revenues.

	Star-film obs.	% Subsidiary	% Sundance	% Star Oscar nominated	Mean compensation	Mean worldwide revenues	Mean compensation / revenues
				Directors (N = 100)			
Age							
18 to 24	9	0%	11%	11%	.	33	.
25 to 29	47	6%	21%	2%	2	108	0%
30 to 34	98	13%	7%	3%	5	186	2%
35 to 39	132	4%	2%	6%	6	163	2%
40 to 44	156	7%	4%	8%	18	181	1%
45 to 49	118	7%	5%	10%	53	201	5%
50 to 54	87	1%	3%	9%	10	137	15%
55 to 59	76	8%	0%	5%	137	129	20%
60 to 64	44	16%	0%	14%	4	104	27%
65 to 69	24	0%	0%	4%	9	88	18%
70 and up	14	14%	0%	14%	.	79	.
All	805	7%	5%	7%	23	158	7%
				Actors (N = 94)			
Age							
5 to 9	10	0%	0%	0%	0	104	0%
10 to 14	26	0%	4%	0%	3	141	7%
15 to 19	53	6%	11%	2%	1	61	1%
20 to 24	201	13%	8%	3%	1	79	3%
25 to 29	396	12%	6%	2%	4	94	8%
30 to 34	434	7%	3%	5%	9	108	12%
35 to 39	399	7%	3%	7%	12	133	11%
40 to 44	289	8%	2%	4%	15	140	16%
45 to 49	193	7%	2%	7%	17	97	17%
50 to 54	139	7%	3%	4%	16	112	17%
55 to 59	87	3%	6%	2%	17	100	15%
60 to 64	35	11%	3%	9%	10	136	15%
65 and up	37	14%	3%	5%	13	110	12%
All	2299	8%	4%	4%	10	111	12%

Table 3.3
OLS Regressions of Log Compensation on Star and Film Characteristics

Sample consists of star-film observations for which compensation data are available from the IMDB. All regressions include dummy variables for film rating (G, PG, PG-13, and R), genre (action, comedy, drama, romance, and thriller), and the job(s) of the star in the film (voice actor, actor, producer, executive producer, director, and writer). *Film count* is the number of prior films the star has made.

Year	0.10
	[0.02] * **
Film count	0.14
	[0.03] * **
Film count ^{2}	−0.0022
	[0.00] * **
Age	0.19
	[0.05] * **
Age ^{2}	−0.0018
	[0.00] * **
Studio	0.63
	[0.23] * **
Subsidiary	−1.19
	[0.29] * **
Sundance	−0.88
	[0.45]*
Constant	−208.0
	[39.01] * **
Director FE	Yes
Actor FE	Yes
Genre FE	Yes
Rating FE	Yes
Job FE	Yes
Observations	495
R-squared	0.57

Table 3.4
Logistic Regressions of Oscar Nominations

The dependent variable in specifications (1) and (2) is a 0/1 indicator for whether the film received an Oscar nomination in the eight major categories: Best Picture, Best Director, Best Actress/Actor, Best Supporting Actress/Actor, and Best Adapted/Original Screenplay. The dependent variable in column (3) indicates a nomination in the Best Director category. The dependent variable in column (4) is a 0/1 indicator for a nomination in one of the four acting categories. *Studio*, *Subsidiary*, and *Sundance* are dummies for film type and *Star director/actor* are dummies indicating the presence of stars in the film. *Dir/Act film count* is the number of films the star director/actor has made before the current film. *Dir/Act sub film count* and *Dir/Act Sun film count* are defined analogously for subsidiary and Sundance films. All specifications include dummy variables for film rating (G, PG, PG-13, and R) and genre (action, comedy, drama, romance, and thriller) as identified in the IMDB. Standard errors are clustered at the studio-year level.

Sample:	All	Star	Star Director	Star Actor
Award:	Any	Any	Director	Acting
	(1)	(2)	(3)	(4)
Studio	1.06			
	[0.24] * **			
Subsidiary	0.40	0.88	1.71	0.46
	[0.15] * **	[0.25] * **	[0.52] * **	[0.40]
Sundance	−0.20	−0.42	−1.39	−1.21
	[0.20]	[0.34]	[1.31]	[0.66]*
Star director	1.47			
	[0.14] * **			
Star actor	0.47			
	[0.13] * **			
Dir age		0.066	0.016	
		[0.01] * **	[0.08]	
Dir age^2		−0.00079	−0.00030	
		[0.00] * **	[0.00]	
Dir film count		0.012	0.058	
		[0.03]	[0.03]*	
Dir subsidiary film count		−0.058	−0.18	
		[0.15]	[0.16]	
Dir Sundance film count		0.24	−0.16	
		[0.17]	[0.26]	
Act age		0.016		0.20
		[0.02]		[0.11]*
Act age^2		−0.00013		−0.0021
		[0.00]		[0.00]
Act film count		0.0053		0.0022
		[0.01]		[0.01]
Act subsidiary film count		0.036		0.014
		[0.07]		[0.06]
Act Sundance film count		0.090		0.24
		[0.09]		[0.07] * **
Constant	−6.30	−4.52	−18.9	−35.0
	[0.42] * **	[0.90] * **	[4.53] * **	[2.12] * **
Genre FE	Yes	Yes	Yes	Yes
Rating FE	Yes	Yes	Yes	Yes
Observations	6611	1849	699	2032
Pseudo R-squared	0.21	0.20	0.17	0.19
Log likelihood	-1273.2	-579.5	-167.1	-324.4

Table 3.5
OLS Regressions of Log Revenues on Film Characteristics

The dependent variable is the log of worldwide revenues. The sample in specification (1) includes all films released between 1980 and 2005 in the Box Office Mojo database. The sample in specification (2) includes all films with a star director or actor. The samples in specifications (3), (4), and (5) include only films with stars and that fall in the mutually-exclusive *Studio*, *Subsidiary*, and *Sundance* categories described in Table 3.1a. Independent variables are as defined in Table 3.4. Standard errors are clustered by studio-year.

	All films	Star films			
		All	Studio	Subsidiary	Sundance
	(1)	(2)	(3)	(4)	(5)
Studio	2.11	1.69			
	[0.14] * **	[0.19] * **			
Subsidiary	−0.74	−0.93			
	[0.10] * **	[0.18] * **			
Sundance	−0.37	−0.81			
	[0.10] * **	[0.25] * **			
Star director	1.32				
	[0.07] * **				
Star actor	1.25				
	[0.07] * **				
Dir age		0.043	0.026	0.081	0.18
		[0.03]	[0.03]	[0.04] **	[0.08] **
Dir age^2		−0.00035	−0.00021	−0.0015	−0.0028
		[0.00]	[0.00]	[0.00]*	[0.00]
Act age		0.020	0.037	0.057	0.091
		[0.01]*	[0.01]***	[0.04]	[0.08]
Act age^2		−0.00027	−0.00057	−0.00049	−0.0011
		[0.00]	[0.00] **	[0.00]	[0.00]
Dir film count		−0.016	−0.019	0.021	0.021
		[0.04]	[0.04]	[0.07]	[0.17]
Dir subsidiary film count		0.11	0.073	0.15	0.80
		[0.09]	[0.11]	[0.23]	[1.02]
Dir Sundance film count		−0.20	−0.051	0.31	−0.28
		[0.28]	[0.28]	[0.47]	[1.17]
Act film count		0.0081	0.0034	0.0067	0.0045
		[0.01]	[0.01]	[0.01]	[0.05]
Act subsidiary film count		0.071	0.13	0.018	0.14
		[0.06]	[0.06] **	[0.09]	[0.22]
Act Sundance film count		0.076	0.11	0.21	−0.13
		[0.10]	[0.10]	[0.16]	[0.36]
Genre FE	Yes	Yes	Yes	Yes	Yes
Rating FE	Yes	Yes	Yes	Yes	Yes
Director FE	No	Yes	Yes	No	No
Actor FE	No	Yes	Yes	No	No
Observations	6499	1845	1461	144	94
R-squared	0.52	0.84	0.89	0.36	0.49

Table 3.6
Artistic Film Preference and Star Value

Pairwise correlations between star career characteristics. *Mean rank* is 101 - the star's mean rank in the *Premier 100* across all of the years she appears on the list (the mean rank measure is thus increasing in star power from 1 to 100). *Adjusted film count* is the star's total film count subtracted by the mean film count within 5-year bins for age in 2005 (adjustments are performed separately for directors and actors). *Revenue fixed-effect* is the star fixed-effect from Specification (2) in Table 3.5.

	% Subsidiary	% Sundance	% Oscar nominated	Mean comp.	Mean rank	Adjusted film count	Mean revenues
			Directors				
% Subsidiary	1.000						
% Sundance	0.286***	1.000					
% Oscar nominated	0.250 **	−0.065	1.000				
Mean compensation	−0.151	−0.137	−0.103	1.000			
Mean rank	0.019	−0.223 **	−0.038	0.283	1.000		
Adjusted film count	−0.155	−0.132	0.035	−0.182	−0.074	1.000	
Mean revenues	−0.200 **	−0.158	0.141	0.627***	0.156	0.142	1.000
Revenue fixed-effect	−0.049	0.089	0.313***	0.216	−0.087	0.097	0.441***
			Actors				
% Subsidiary	1.000						
% Sundance	0.315***	1.000					
% Oscar nominated	0.058	0.133	1.000				
Mean compensation	−0.289 **	−0.303***	0.124	1.000			
Mean rank	−0.055	−0.246 **	0.023	0.504***	1.000		
Adjusted film count	−0.012	−0.101	0.006	0.209*	0.165	1.000	
Mean revenues	0.018	0.065	−0.009	0.491***	0.511***	0.045	1.000
Revenue fixed-effect	−0.280***	−0.046	0.031	0.257 **	−0.018	−0.086	0.303***

Appendix A
Studios and Subsidiaries

The table shows the seven major studios and their subsidiaries, with artistic subsidiaries marked by *. Ownership changes consisted of the following: ** MGM and all subsidiaries were purchased by Sony in 2005; American International Films, Filmways, Orion Pictures, and Orion Pictures Classics were owned by Warner Bros. from 1980 until 1989 and by MGM between 1989 and 2005. When ownership changes occur, films released during the year of the change are attributed to the previous owner, and films released starting the year after the change are attributed to the new owner.

Studio	Subsidiaries
Disney	Buena Vista Caravan Pictures Dimension Hollywood Pictures Miramax * Touchstone Pictures
Universal	Focus Features * Good Machine * Gramercy * October October Classics * Polygram Rogue Pictures USA Films *
20th Century Fox	Blue Sky Fox Intl Classics * Fox Searchlight *
MGM**	American International Pictures Filmways Orion Classics * Orion Pictures Samuel Goldwyn * United Artists Columbia Screen gems Sony Classics * Sony Repertory * TriStar
Sony Warner Bros.	Castle Rock Fine Line * New Line Warner Independent *
Paramount	Paramount Classics * Republic

Bibliography

J. M. Abowd, J. Haltiwanger, R. Jarmin, J. Lane, P. Lengermann, K. McCue, K. McKinney, and K. Sandusky. The Relation Among Human Capital, Productivity, and Market Value. In M. M. Blair and T. A. Kochan, editors, *Measuring capital in the new economy*, pages 153–198. 2005.

D. Acemoglu and J. S. Pischke. Beyond Becker: Training in Imperfect Labour Markets. *The Economic Journal*, 109 (453):112–142, 1999.

V. V. Acharya, S. T. Bharath, and A. Srinivasan. Does Industry-wide Distress Affect Defaulted Firms? Evidence from Creditor Recoveries. *Journal of Financial Economics*, 85(3):787–821, 2007.

S. Albert. Movie stars and the distribution of financially successful films in the motion picture industry. *Journal of Cultural Economics*, 22(4):249–270, 1998.

M. J. Alderson and B. L. Betker. Liquidation Costs and Capital Structure. *Journal of Financial Economics*, 39(1): 45–69, 1995.

E. I. Altman. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*, pages 589–609, 1968.

P. Asquith, R. Gertner, and D. Scharfstein. Anatomy of Financial Distress: An Examination of Junk-Bond Issuers. *Quarterly Journal of Economics*, 109(3):625–658, 1994.

D. H. Autor. Why do Temporary Help Firms Provide Free General Skills Training? *Quarterly Journal of Economics*, 116(4):1409–1448, 2001.

D. H. Autor, F. Levy, and R. J. Murnane. The Skill Content of Recent Technological Change: An Empirical Exploration. *Quarterly Journal of Economics*, 118(4):1279–1333, 2003.

D. G. Baird and R. K. Rasmussen. The End of Bankruptcy. *Stanford Law Review*, 55:751–790, 2002.

L. J. Bassi, B. Lev, J. Low, D. P. McMurrer, and G. A. Seisfeld. Measuring Corporate Investments in Human Capital. In M. M. Blair and T. A. Kochan, editors, *The New Relationship: Human Capital in the American Corporation*, pages 334–381. 2000.

S. Basuroy, S. Chatterjee, and S.A. Ravid. How critical are critical reviews? The box office effects of film critics, star power, and budgets. *Journal of Marketing*, pages 103–117, 2003.

G. S. Becker. Investment in Human Capital: A Theoretical Analysis. *The Journal of Political Economy*, 70(S5):9–49, 1962.

G. S. Becker. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* National Bureau of Economic Research, 1964.

Y. Ben-Porath. The Production of Human Capital and the Life Cycle of Earnings. *The Journal of Political Economy*, pages 352–365, 1967.

J. Berk, R. Stanton, and J. Zechner. Human Capital, Bankruptcy and Capital Structure. *NBER Working Paper*, 2007.

E. Berkovitch, R. Israel, and J. F. Zender. Optimal Bankruptcy Law and Firm-Specific Investments. *European Economic Review*, 41(3-5):487–497, 1997.

S. T. Bharath, V. Panchapagesan, and I. M. Werner. The Changing Nature of Chapter 11. *Mimeo, University of Michigan*, 2007.

P. Biskind. *Down and Dirty Pictures: Miramax, Sundance, and the Rise of Independent Film.* Simon & Schuster, 2004.

M. Bradley and M. Rosenzweig. The Untenable Case for Chapter 11. *Yale Law Journal*, 101(5):1043–1095, 1992.

D.T. Brown, C.M. James, and R.M. Mooradian. Asset Sales by Financially Distressed Firms. *Journal of Corporate Finance*, 1(2):233–257, 1994.

A. De Vany and W.D. Walls. Uncertainty in the Movie Industry: Does Star Power Reduce the Terror of the Box Office? *Journal of Cultural Economics*, 23(4):285–318, 1999.

A. De Vany and W.D. Walls. Does Hollywood Make Too Many R-Rated Movies? Risk, Stochastic Dominance, and the Illusion of Expectation. *The Journal of Business*, 75(3):425–451, 2002a.

A. De Vany and W.D. Walls. Movie Stars, Big Budgets, and Wide Releases: Empirical Analysis of the Blockbuster Strategy. 2002b.

A.S. De Vany. *Hollywood Economics.* Routledge, 2004.

D. K. Denis and K. J. Rodgers. Chapter 11: Duration, Outcome, and Post-Reorganization Performance. *Journal of Financial and Quantitative Analysis*, 42(1):101–118, 2007.

A. Elseberse. The Power of Stars: Creative Talent and The Success OF Entertainment Products. 2006.

H. S. Farber. Mobility and Stability. *Handbook of Labor Economics*, 3:2441–2483, 1999.

H. S. Farber. *Job Loss and the Decline in Job Security in the United States.* 2007.

C.E. Fee. The Costs of Outside Equity Control: Evidence from Motion Picture Financing Decisions. *The Journal of Business*, 75(4):681–711, 2002.

C. Frydman. Rising through the ranks. the evolution of the market for corporate executives, 1936-2003. *Unpublished manuscript. Cambridge, MA: Harvard University*, 2005.

S.C. Gilson. Bankruptcy, Boards, Banks, and Bondholders: Evidence on Changes in Corporate Ownership and Control When Firms Default. *Journal of Financial Economics*, 27:355–388, 1990.

S.C. Gilson, K. John, and L. Lang. Troubled Debt Restructurings: an Empirical Study of Private Reorganization of Firms in Default. *Journal of Financial Economics*, 27:315–353, 1990.

W. Goldman. *The Big Picture: Who Killed Hollywood? and Other Essays.* Applause, 2000.

R.E. Hall and E. Lazear. The Excess Sensitivity of Quits and Layoffs to Demand. *Journal of Labor Economics*, 2(2): 233–257, 1984.

M. Hashimoto. Firm-Specific Human Capital as a Shared Investment. *American Economic Review*, 71(3):475–482, 1981.

E. S. Hotchkiss. Post-Bankruptcy Performance and Management Turnover. *Journal of Finance*, 50(1):3–21, 1995.

L. S. Jacobson, R. J. LaLonde, and D. G. Sullivan. Earnings Losses of Displaced Workers. *The American Economic Review*, pages 685–709, 1993.

D. A. Jaeger and A. H. Stevens. Is Job Stability in the United States Falling? Reconciling Trends in the Current Population Survey and Panel Study of Income Dynamics. *Journal of Labor Economics*, 17(S4):1–28, 1999.

P.B. Jaggia and A.V. Thakor. Firm-specific human capital and optimal capital structure. *International Economic Review*, 35(2):283–308, 1994.

K. John, S.A. Ravid, and J. Sunder. The Role of Termination in Employment Contracts: Theory and Evidence from Film Directors Careers. 2002.

D. W. Jorgenson, M. S. Ho, and K. J. Stiroh. Growth of US Industries and Investments in Information Technology and Higher Education. *Economic Systems Research*, 15(3):279–325, 2003.

B. Jovanovic. Job Matching and the Theory of Turnover. *The Journal of Political Economy*, 87(5):972–990, 1979.

A. Kalay, R. Singhal, and E. Tashjian. Is Chapter 11 Costly? *Journal of Financial Economics*, 84(3):772–796, 2007.

L. F. Katz and K. M. Murphy. Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *Quarterly Journal of Economics*, 107(1):35–78, 1992.

L. F. Katz and L. H. Summers. Industry Rents: Evidence and Implications. *Brookings Papers on Economic Activity. Microeconomics*, pages 209–290, 1989.

M. L. Lemmon, Y. Y. Ma, and E. Tashjian. Survival of the Fittest? Financial and Economic Distress and Restructuring Outcomes in Chapter 11. *Unpublished working paper*, 2009.

B.R. Litman and L.S. Kohl. Predicting Financial Success of Motion Pictures: The'80s Experience. *Journal of Media Economics*, 2(2):35–50, 1989.

Colin McCabe. Interview with Peter Biskind, August 20, 2004. URL `http://www.bfi.org.uk/features/interviews/biskind.htmlindependent`.

J. Mincer. Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2. 1974.

J. Mottram. *The Sundance kids: how the mavericks took back Hollywood.* Faber & Faber, 2006.

K.J. Murphy. Executive compensation. *Handbook of Labor Economics*, 3:2485–2563, 1999.

D. O. Parsons. Specific Human Capital: An Application to Quit Rates and Layoff Rates. *The Journal of Political Economy*, pages 1120–1143, 1972.

T. C. Pulvino. Effects of Bankruptcy Court Protection on Asset Sales. *Journal of Financial Economics*, 52(2):151–186, 1999.

R.G. Rajan and J. Wulf. Are perks purely managerial excess?$. *Journal of Financial Economics*, 79:1–33, 2006.

S.A. Ravid. Information, Blockbusters, and Stars: A Study of the Film Industry. *The Journal of Business*, 72(4): 463–492, 1999.

S.A. Ravid and S. Basuroy. Managerial Objectives, the R-Rating Puzzle, and the Production of Violent Films. *Journal of Business*, 77(2):155–192, 2004.

S. Rosen. The economics of superstars. *The American Economic Review*, 71(5):845–858, 1981.

A. Shleifer and R.W. Vishny. Liquidation Values and Debt Capacity: A Market Equilibrium Approach. *Journal Of Finance*, 47:1343–1366, 1992.

S. Stern. Do scientists pay to be scientists? *Management Science*, pages 835–853, 2004.

M. Tervio. Mediocrity in Talent Markets. 2003.

R. Topel. Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority. *Journal of Political Economy*, pages 145–176, 1991.

C. Vachon and A. Bunn. *A Killer Life: How an Independent Film Producer Survives Deals and Disasters in Hollywood and Beyond*. Simon & Schuster, 2006.

S. Waxman. *Rebels on the backlot: six maverick directors and how they conquered the Hollywood studio system*. Harper Entertainment, 2005.

L. A. Weiss. Bankruptcy Resolution: Direct Costs and Violation of Priority of Claims. *Journal of Financial Economics*, 27:285–314, 1990.

I. Welch. Common Flaws in Empirical Capital Structure Research. *Working Paper*, 2008.

D. Yermack. Flights of Fancy: Corporate Jets, CEO Perquisites, and Inferior Shareholder Returns. *Unpublished working paper (New York University)*, 2004.

L. Zingales. In Search of New Foundations. *Journal of Finance*, pages 1623–1653, 2000.